INVESTIGATING A RELATIONSHIP OF A PERCEPTION OF AGENCY IN TASK BASED DISCOURSE AND CHANGE IN CONCEPTS:

A Practitioner Research in Education

Lubna Khalid KIDWAI

Submitted for the Degree of
Doctor of Philosophy

Faculty of Social Sciences
University of Bradford

2015

Investigating a Relationship of a Perception of Agency in Task Based Discourse and Change in Concepts: A Practitioner Research in Education

Lubna Khalid Kidwai

ABSTRACT

This research is a practitioner inquiry into change in concepts in contexts of task based discourse in teacher education classrooms. The particular concepts selected for study are epistemic beliefs about knowledge in history. A hypothesis that there is a relationship between a perception of agency in task based discourse and change in concepts is proposed for research on the basis of a review of Kuhn's arguments for paradigm change and those of others on the problems of convergence of meaning between paradigms.[1] Four of five quasi experimental studies conducted find tentative support for the alternative hypothesis and demonstrate how the relationship could be tested in the field.

Change in epistemic beliefs from pre-test to post-test is significant in both groups, Control and Experimental, and consistently more and better change with moderate effect sizes is seen in the Experimental groups. A matched pedagogical method, experience of a range of suitable examples, Skemp (1971) was provided to both Control and Experimental groups and this is found to be effective in developing beliefs. The experimental construct, a perception of agency in task based discourse was enhanced in Experimental groups alone, which is used to explain the difference in change.

A rubric, the Categories of Beliefs about Knowledge in History, CBKH, is developed and employed to examine responses to open ended questions. Likert Scales are also used and conclusions are drawn with a final quantitative analysis of data. Excerpts from a focus group discussion illustrate the process.

Keywords: Conceptual Change, Task Based Discourse, Perception of Agency, Epistemic Beliefs, Teacher Education

[1] Among others, Schön (2002), Newman (1999) and Russell (1996).

ACKNOWLEDGEMENTS

To begin, I thank Allah Subhanahu wa Talla for His Blessings.

For the development of my ideas I would like to thank Mr Ian Fouweather of Graduate School, University of Bradford, who, especially with his deep understanding, wit and willingness to respond to all emails, has made my efforts to learn practical philosophy amusing and challenging. Dr Liliana Maggioni from the University of Maryland was the expert in the field, whose generous and untiring support made the development of the Categories of Beliefs about Knowledge in History, the CBKH Scale, and the assessment of responses possible. She was always ready to read the documents I sent her and respond to all my questions. An academic in America helping a student in Pakistan doing research from Bradford University UK, is only possible because people are human and love to help others, and of course love a challenge!

Mrs Nadira Mirza, Dean of SLED, was the University for me, the extra mural student. From discussing issues in education, visiting schools, carefully reading documents, making study and stay arrangements, and even providing a room upstairs, her support was always available. Thank you Nadira!

From the depths of my heart, I would like to thank my husband Khalid Ahmed Kidwai for his constant, untiring, intellectual and emotional support in my mad plan for a PhD. This is not to underplay Khalid's help in proof reading and organization of the thousands of words, chapters, sections, tables, etcetera that took days and nights of slog. Thank you, Khalid. Thank you Samaya, for your presence by my side and for Misha who at 8 months wanted to take over the key board! Thank you Yasser, my son, you with your quizzical, amused encouragement and bright ideas that lightened the hard work and kept me going.

To all my supervisors and advisors, Professor Geoff Layer, Mrs Nadira Mirza, Doctors Sheeran and Paul Jennings, Ms Mercedes Webb, Professor Yunas Samad, Mr Ian Fouweather, who bore with the scale and complexity into which I had ventured, with patience and encouragement, I want to say thank you. This research would never have come about if Professor Layer and Mrs Nadira Mirza

had not found something of value in my proposal and agreed to take on a hopeful candidate for extra mural study from Pakistan. Dr Ivan Reid supported me in developing the research design, offering journals to read and asking questions that were hard to answer. Not to mention the trip to the railway factory! It was Graduate School and the modules of study that developed my understanding of research and the philosophies that underpin it and for this again I need to thank Ian Fouweather who was there from the start. To Dr Samad, who was willing to accept me as a student, and advised, pushed and supported me to the finish, thank you!

For Donna Irwing, Maria Introwicz, Rebecca Woods, who kept the support going from J.B.Priestly Library all these years, Thank you!

To my beloved parents, my mother in law and father in law who did not live to see this, I wish you were here; my family, all brothers and sisters, family in England, thank you, love you all, and I promise I won't do it again!

To Dr Anila Kamal,[2] Professor and Chairperson of the Department of Psychology, Quaid-e-Azam University Islamabad, with her time, generous and informed guidance in discussing issues of design and analysis; Dr Razia Sultana[3] and students of the Department of History, Quaid-e-Azam University, for helping out in the pilot and inter-rater exercise. Dr Faraz, a colleague, Dr Junaid of Karachi University for their valuable advice in statistical analysis, Mr Sabah Qaiser of Karachi University for allowing me to think I was in charge of SPSS when I outlined the relationships that I sought and the things that did not make sense. I want to thank Mr Asad Nizam for his constant support in IT, Mr Zil Muhammad who kept the papers and the data and filed the documents, fixed the computer glitches, and produced the records that I could not find when I needed them. This certainly could not have been possible without him. To all those who kept the fort while mom was 'busy', thank you.

In the end I want to thank all my colleagues in the Beaconhouse School System Pakistan, and all those children and teachers who were part of my journey

[2] Dr Anila Kamal was a keynote speaker at ICOPE 6th International Conference on Psychology Education

[3] Dr Razia Sultana is currently Vice Chancellor Shaheed Benazir Bhutto Women's University Peshawar Pakistan

in academics. I believe the strength of my study at this late stage in my career is due to my rich experience in teaching which I began as a teenager when I taught my siblings, and the little boy whose mom worked for us who said, ‘Sister I understand what is written but can you tell me what is a Mary?’

Glossary

The meanings of key terms are discussed in depth in the pages that follow while a simple definition or reference is provided below.

- A Perception of Agency: An enhanced perception of agency to reflect upon one's prior knowledge, to question, argue, agree or disagree with group views, progress at one's pace, is facilitated within the task-based group discourse in the quasi experiments conducted for this research. The manner in which a perception of agency is operationalised as an intervention is described in the research methodology. The experience is provided to experimental groups exclusively and not to control groups.
- Task Based Group Discourse: This is a term coined for a group discussion in the context of an active learning task. The term encompasses the contexts, actions and words required for concept development.
- Change in Concepts: What are concepts and change within these is not open for simple definition as is argued in Section 2.2.2. The concepts under consideration in this research are epistemic beliefs about knowledge in history. Change is considered as a move on a scale from naive to sophisticated beliefs.
- Experience of a Range of Suitable Examples: This term is derived from Skemp (1971, p.32). A Psychology of Learning Mathematics. There is a small amendment from Skemp's term 'collection' to 'range'. The meaning here is near literal with each word of consequence. See Section 2.2.11.

List of Acronyms

- BM: Boscolo & Mason
- BLTHQ: Beliefs about Learning and Teaching History Questionnaire
- EBI: Epistemic Beliefs Inventory
- CBKH: Categories of Beliefs about Knowledge in History
- CIE: Cambridge International Examinations
- O' Level: Ordinary Level Examination
- C: Control
- E: Experimental

Citation and Referencing in the Dissertation

- Harvard Referencing is used throughout with Endnote 7 software.
- The Reference List has all sources cited in the document.
- Following Harvard Guide to Using Sources (2015) on exceptions to citation being material that is common knowledge and, therefore not attributable to one source, Kuhn's theory of paradigm change (1962), the work of Piaget (1928) and Vygotsky (1934) is not written with a citation each time. As Kuhn and Kuhn's theory is referred to multiple times in this document a reference is added the first time and in places where a particular item is quoted but not each time.

TABLE OF CONTENTS

Title
Abstract
Acknowledgements i
Glossary iv
List of Acronyms v
Citation and Referencing in the Dissertation v
Table of Contents vi
List of Tables xvii
List of Figures xx
CHAPTER 1 INTRODUCTION 1
1.1 Introduction to the Problem 1
1.2 Theoretical Background in Teacher Education 3
1.3 Competing Explanations of Conceptual Change 4
1.4 The Theoretical Framework: An Argument for Agency 6
1.5 Pedagogical Method for the Experiments 8
1.6 Research Aim 9
1.7 Study Question 9
1.8 The Research Hypothesis 9
1.9 Key Objectives of Research 9
1.10 Purpose of Study 10
1.11 Benefits of the Study 11
1.12 Importance and Scope of the Study 14
CHAPTER 2 THE REVIEW OF LITERATURE 18
2.1 Introduction 18
2.2 Concepts, Epistemic Beliefs? 19
2.2.1 Perspectives on Conceptual Change 21
2.2.2 Conceptual Change in Various Domains 25

2.2.3 The Use of Kuhn's Theory (1970) and Posner et al's Model (1982) in Conceptual Change Theory Building 27
2.2.4 Understanding Conceptual Change 30
2.2.5 Intentional Conceptual Change 33
2.2.6 Adult Learning in Groups 35
2.2.7 The Argument for Agency 36
2.2.8 Schön's Epistemology 39
2.2.9 Agency and Mental Development from a Perspective of Cognitive Psychology 41
2.2.10 Reflection in the Process of Conceptual Change 44
2.2.11 Reflection on My Own Learning and Epistemic Belief Change 48
2.2.12 Skemp's Principles for the Formation of Concepts 49
2.2.13 Reasons for a Focus on Epistemic Beliefs and Beliefs About Knowledge in History 49
2.2.14 The Importance of Research into the Development of Concepts About Knowledge in History 51
2.2.15 Concerns and Purposes of Teaching History 53
2.3 Review of Literature for Research Methodology 53
2.3.1 Arguments on Paradigms and Mixed Methods 53
2.3.2 Effect Sizes 62
2.4 Arguments for the Design of the Quasi Experiment and Validity of Findings in Quasi-experimentation with Reference to this Research 63
2.5 Truth Claims in Quasi Experimental Research 64
2.5.1 The Assignment Process 65
2.5.2 Validity 67
2.5.3 Fuzzy Plausibility 68
2.5.4 Types of Designs 69
2.5.5 Selection Bias 70
2.5.6 Generalized Causal Inference 71
2.5.7 Possibilities of Randomization and Standardization 72

2.5.8 Cause in Quasi Experimentation 73
2.5.9 Alternative Explanations in Causal Inference 74
2.5.10 Thoughts on the Review of Literature on Quasi-Experimentation 75
2.6 Theoretical Background of Measures and the CBKH Rubric 76
2.6.1 Theoretical Models of Epistemic Beliefs that Inform the Development of the BLTHQ, Beliefs about Learning and Teaching History Questionnaire and the CBKH, Categories of Beliefs about Knowledge in History 77
2.6.2 The Model of Copier, Borrower, and Criterialist Categories of Epistemic Beliefs of Knowledge in History 81
2.6.3 Are Epistemic Beliefs Domain General or Domain Specific Beliefs? 82
2.6.4 Does Culture Have an Effect on Epistemic Beliefs? 84
2.6.5 Epistemic Beliefs in the Domain of Knowledge in History and How These Differ from Domain General Epistemic Beliefs 85
2.6.6 Development and Validation of the CBKH, the Categories of Beliefs about Knowledge in History 87
2.6.6.1 Background for the Development of the CBKH 87
2.6.6.2 Whose Beliefs? 88
2.6.6.3 Developments from Maggioni et al Model to the CBKH 89
2.6.7 Question 1: Why is it Necessary to Unravel Elements of Belief in Knowledge in History? 92
2.6.8 Question 2: Is the Criterialist Just a More Efficient Objectivist? 96
2.6.9 Question 3: Is Awareness of the Acknowledged Presence of the Historian in the Text a Necessary Aspect of Sophistication for the Reader of History? 99
CHAPTER 3 THE RESEARCH METHODOLOGY 102
3.1 Introduction 102
3.2 Aim of Research 105

3.3 Study Question 105
3.4 The Research Hypothesis 105
3.5 Key Objectives of Research 105
3.6 Specific Objectives for Conduct of Research and Analysis 106
3.7 Objectives of Data Collection 106
3.8 Objectives of Analysis 106
3.9 Features and Steps for Analysis 107
3.10 Objectives for Discussion and Conclusion Based upon Qualitative and Quantitative Data and its Analysis 108
3.11 Research Facts 109
3.11.1 Design of Quasi Experiment 109
3.11.2 Sample Size 109
3.11.3 Dependent Variable 110
3.11.4 Independent Variable 110
3.12 Features of Mixing of Methods Used in this Research 111
3.12.1 Quantitative Data 111
3.12.2 Qualitative Data 112
3.13 Description of Research Methods 112
3.14 Answering Research Questions Using Mixed Methods 113
3.15 Background 114
3.16 Locating the Study 116
3.17 Procedures 117
3.18 Research Constraints and Opportunities 118
3.19 The Design and Data Collection (Appendix-A6) 119
3.20 The Intervention and Standardization (Appendix-A6) 121
3.20.1 During the Session the Following Steps were Taken 122
3.20.2 Experience in the Control Groups with Reference to the Experimental Treatment 125
3.20.3 How was the Experience Different in the Control Group? 125
3.21 Materials Used for Learning and Teaching 126

3.22 Topics and Examples of Multiple Perspectives and Source Material Used in Workshops: 127
3.23 The Sample and Control 127
3.24 Instrument Effects 131
3.25 Focus Group as a 'A Feedback Session' 132
3.26 Data Collection and Analysis 133
3.27 Reliability 134
3.28 Inter Rater Reliability and Internal Evaluation 134
3.29 Ethical Issues 134
3.30 Some Steps That Were Taken Are 135
3.31 Development of Knowledge and Skills 137
3.32 Description of the Quasi Experiments 137
3.32.1 Introduction 137
3.32.2 Truth Claims in Quasi Experimental Research 138
3.32.3 Ambiguous Temporal Precedence 140
3.32.4 Selection and Assignment for Control 140
3.32.5 Assignment to Control and Experimental Groups 142
3.32.6 Differential Attrition 142
3.32.7 Access and Social Situation 143
3.33 Alternative Explanations 145
3.33.1 Prior Study 145
3.33.2 Knowledge of the Treatment 145
3.33.3 History and Maturation 146
3.33.4 Threat of Testing 147
3.33.5 Resentful Demoralization 148
3.33.6 Threat of Difference in Experience with Different Tutors 149
3.33.7 Size of Sample 149
3.33.8 When Treatment is Non-Obtrusive or Comparison Becomes Difficult 150
3.33.9 Non Availability of Baseline Assessment 150
3.33.10 Confounding Factors 151

3.33.11 Construct Validity 151
3.33.11.1 Inadequate Pre-Operational Explication of Constructs 151
3.33.11.2 Explication of Treatment 152
3.33.11.3 Variables in the Sample and Settings and Tutor 153
3.34 Variables in the Treatment 156
3.34.1 Slip Between Intention and Action in Classroom Practice 156
3.35 The Measures 159
3.35.1 A Measure of a Perception of Agency? 159
3.35.2 Questions to Consider in Instrument Selection 160
3.35.2.1 The Open Ended Boscolo and Mason Questions: 161
3.35.3 Some Considerations 161
3.35.4 Measure of Beliefs about Knowledge in History 161
3.35.5 The Epistemic Beliefs Inventory 162
3.36 The CBKH Category Scale 163
3.36.1 Introduction to the Proposed Scoring Rubric: Categories of Beliefs about Knowledge in History the CBKH 163
3.36.2 Selecting the Instrument of Data Collection 165
3.36.3 Categories of Beliefs about Knowledge in History (CBKH) 166
3.36.4 Qualitative Analysis of Responses 168
3.36.5 Category Scale Options with Weightage 170
3.36.6 Selection of a Range of Responses for Each Question 186
3.37 Validation of Assessment 193
3.37.1 Assessment of Reliability of Marking of Responses to the Boscolo and Mason Questions 193
3.37.2 Inter Rater Agreement for Validation of Category Table Inter Rater Exercise 195
3.37.3 The Raters 195
3.37.4 Selecting a Sample for Inter Rater Analysis 195
3.37.5 Process 196
3.37.6 Tasks for Raters 196

3.37.7 Questions for the Inter Rater Moderation 197
3.37.8 Criteria for Inter Rater Marking 197
3.37.9 Discussion on the Process of Inter Rater Agreement for Validation of the Category Table 197
3.37.10 Concepts and Skills Required of the Raters 198
3.38 Comments on the Exercise 200
3.39 Issues 201
3.40 Summary of Chapter 203
CHAPTER 4 RESULTS AND DISCUSSION 205
4.1 Introduction 205
4.2 Brief Outline of Objectives and Analysis Decisions in the Context of the Research 208
4.2.1 Missing Data 210
4.3 Analysis of Data from Responses to the Boscolo & Mason Questionnaire 211
4.3.1 Brief Introduction to the Analysis of Data from Responses to the Boscolo & Mason Questionnaire using the CBKH Category Scale 211
4.3.2 Questions for Analysis 211
4.3.3 Statistical Tests Conducted in the Analysis of the BM Data 213
4.3.4 Test of Reliability 215
4.3.5 Results of Final Internal Evaluation of Qualitative Assessment and Scoring of Responses to Boscolo & Mason Questions Using the CBKH Category Scale (Appendix-A7) 215
4.4 Statistical Analysis of Response of Participants at Levels of Epistemic Beliefs in Control and Experimental Groups 215
4.4.1 Comparison of Pre and Post Responses Using Chi Square Test of Independence 215
4.4.2 Odds Ratio for Change and No Change in Epistemic Beliefs 221
4.4.3 Analysis of Pre and Post Difference in Control and Experimental Group Using Paired t Test 222

4.4.4 Analysis of Pre-Test and Post-Test Difference in Five Studies for the Control and Experimental Groups 223
4.4.5 Estimating Intervention Effect in the Experimental Group in Meta-Analysis of BM Data 223
4.5 Fine Grain Analysis of Change in Respondents Stance 227
4.5.1 Fine Grain Analysis of Classes, No Change, Degree and Category Change 227
4.5.2 Results of Mann Whitney Test of Classes of Change, No Change, Degree Change, and Category Change 230
4.5.3 Fine Grain Analysis of Low and High Stance for Control and Experimental Groups in Overall Combined Data 231
4.5.4 Fine Grain Analysis of Low and High Stance in the Five Study Groups 234
4.5.5 Difference-In-Difference Estimator after Excluding Pansy 236
4.6 Summary of Analysis of Data from Responses to BM (Boscolo and Mason) Questions 237
4.6.1 Reliability of the Instrument 237
4.6.2 Frequency Distribution Table with Chi-Square Test 238
4.6.3 Odds Ratio for Change and No Change 238
4.6.4 Pre-Test and Post-Test Difference in Both Groups Using a Paired t Test 238
4.6.5 Pre-Test and Post-Test Difference in 5 Studies Using Wilcoxen Rank Sum Test 238
4.6.6 Estimating Intervention Effect Using Difference-In-Difference Estimator 239
4.6.7 Fine Grain Analysis of Change in Stance (No Change, Degree Change and Category Change) with a Mann Whitney Test 239
4.6.8 Analysis of Low and High Stance in Overall Control and Experimental Groups 239

4.6.9 Fine Grain Analysis of Low and High Stance in 5 Independent Studies Using a D.I.D Estimator 239
4.7 Analysis of Data from the Likert Scale, the Belief about Learning Teacher History Questionnaire (BLTHQ) 240
4.7.1 Questions for Analysis of the BLTHQ Data 240
4.7.2 Reliability of the Instrument 240
4.7.3 Arrangement of Data 241
4.7.3.1 Change as Observed in Responses to the BLTHQ Items is Expressed in the Following Arrangement: 241
4.7.3.2 Results and Discussion of Analysis on Cross Tables and Chi Square 242
4.7.3.3 Quantifying Difference of the Difference in Change as seen on Cross Tables 243
4.7.4 List of Statistical Analysis of the BLTHQ Items to Identify and Analyze Change 246
4.7.4.1 Cross Table and Chi Square 246
4.7.4.2 Analysis of Stance from Cross Tables with Percentages at Agreement and Disagreement 247
4.7.4.3 Analysis of the Difference of the Difference 247
4.7.4.4 Wilcoxen Rank Sum Test 248
4.7.4.5 Odds Ratio 249
4.7.4.6 Analysis of Stance 249
4.7.4.7 Exploratory Factor Analysis 249
4.7.4.8 Weighted Score and Consistency Score of Epistemic Beliefs 250
4.7.5 Description of Analysis of Change 251
4.7.5.1 Cross Table of BLTHQ Items 251
4.7.5.2 Comparison of Percentage Change and Chi Square Value for Control and Experimental Groups to Identify a Pattern in the Data 267

4.7.6 Effect Size to Measure the Effectiveness of Intervention 269
4.7.7 Analysis of Selection of the Neutral Option 270
4.7.8 Analysis of Individual Change from Pre-Test to Post-Test 271
4.7.9 Analysis of Pre-Test to Post-Test Difference in Control and Experimental Groups 276
4.7.9.1 Wilcoxen Rank Sum Test 276
4.7.10 Analysis of Change and No Change in Control and Experimental Groups 279
4.7.11 Exploratory Factor Analysis of the BLTHQ Items (Appendix-C) .. 282
4.7.12 Factor Analysis of the BLTHQ Combined Data (166 Respondents) 287
4.7.13 Weighted Scores and Consistency Scores of Epistemic Stance for All Participants 289
4.7.14 Weighted Averages and Consistency Score for the Control and Experimental Group 298
4.7.15 Paired t Test of Pre-Test and Post-Test Consistency Scores in Control and Experimental Groups 300
4.8 Epistemic Beliefs Inventory 301
4.8.1 Summary and Discussion of Analysis of Data of the Third Measure, the Epistemic Beliefs Inventory 301
4.8.2 Factor Analysis of the Overall EBI Data 306
4.9 Summary of Results of All Three Measures 307
CHAPTER 5 CONCLUSION 310
5.1 Introduction 310
5.2 The Argument 312
5.3 Reflection and Language Games 314
5.4 Effects of Intervention One, Experience of a Range of Suitable Examples on Change in Concepts 315
5.5 Support in the Data for the Hypothesis of a Relationship of a Perception of Agency in Task Based Discourse with Change in Concepts 316

5.6 Limitations and Strengths 321
5.7 Is the Relationship of a Perception of Agency and Change in Concepts Possible to Research? 324
5.8 Quality of Change as seen in the Three Measures and the Use of the CBKH Categories 326
5.9 Qualitative and Quantitative Analysis and Interpretations 332
5.10 Contribution to Knowledge 335
5.11 Alternative Explanations of Cause 338
5.12 Participants Giving Reasons for Change 340
5.13 Final Thoughts 340

APPENDIX A1 Case Study 1 and Case Study 2 342
APPENDIX-A2 An Information Sheet 344
APPENDIX-A2a The History Workshop Certificate 348
APPENDIX-A2b History Workshops: Samples in All 5 Studies and in Total for Meta-Analysis 349
APPENDIX-A3 Questions Adapted from those used by Boscolo and Mason 351
APPENDIX-A4 Beliefs About Learning and Teaching History 353
APPENDIX-A5 Epistemic Beliefs Inventory 357
APPENDIX-A6 Explaining the Research Design and Intervention 361
APPENDIX-A7 Evaluation of Qualitative Analysis of Boscolo and Mason's Open Ended Questions to the Rubric, Categories of Beliefs About Knowledge in History (CBKH) .. 364
APPENDIX-B Tables 1B to 27B 367
APPENDIX-C Discussion of Exploratory Factor Analysis 388

REFERENCES 392

LIST OF TABLES

Table 3.1: The Categories of Beliefs about Knowledge in History –CBKH .. 169
Table 3.2: Degrees of Pre- Post Change 170
Table 3.3: Degree of Consistency in Responses 170
Table 3.4: First Category Scale with Weighted Scores 171
Table 3.5: The CBKH Categories and their Elements with Reference to Evidence in the Responses..... 172
Table 3.6: The CBKH Categories and their Elements with Reference to Evidence in the Responses..... 176
Table 3.7: 'Sophisticated Integrator' of the Subjective and Objective Dimensions of Knowing..... 184
Table 3.8: Exemplars from Responses to Boscolo and Mason Questions... 187
Table 3.9: Pre-Test and Post-Test Assessment of Raters' Concepts about Knowledge in History (Instrument: responses to the Boscolo & Mason questions)..... 199
Table 3.10: Inter Rater Assessment of Responses Comparison of Scores by Researcher and 3 Raters 202
Table 4.1: Categories and Corresponding Holistic Scores of Epistemic Beliefs in the CBKH..... 214
Table 4.2: Frequency Distributions of Participants at Levels of Epistemic Beliefs in Control and Experimental Groups, Pre-Test and Post-Test..... 217
Table 4.3: Change and No Change Responses in Control and Experimental Group 222
Table 4.4: Analysis No Change, Degree Change, Category Change in Control and Experimental Groups..... 228
Table 4.5: Analysis No Change, Degree Change, Category Change in Control and Experimental Groups 233
Table 4.6: Analysis of Low and High Stance in each City's Study Group for Control and Experimental Groups..... 235
Table 4.7: Difference-In-Difference (D.I.D) Estimator - City Wise 236

Table 4.8: Students Who are Good in Memorization Learn History Quickly 252
Table 4.9: Corroborating Evidence and Identifying Sources are Important Learning Strategies in History but Only After Mastering the Basic Facts 254
Table 4.10: Teachers Need to Avoid Giving Students Conflicting Sources, since it Makes Historical Investigation Impossible 256
Table 4.11: Knowledge of The Historical Method is Fundamental for Historians and Students Alike 258
Table 4.12: Comparing Sources and Looking for Author Subtext are Essential Components of The Process of Learning History 260
Table 4.13: Students need to be taught to Deal with Conflicting Evidence 262
Table 4.14: It is Fundamental that Students are Taught to Support their Reasoning with Evidence and Ask that History Textbook Authors do so Also 264
Table 4.15: Students need to be Aware that History is Essentially a Matter of Interpretation 266
Table 4.16: Comparison of Chi Square Value and Percentage Difference in Control and Experimental Groups 267
Table 4.17: Neutral Option Analysis in the Control and Experimental Group 271
Table 4.18: Analysis of Change in the Control Group 274
Table 4.19: Analysis of Change in the Experimental Group 274
Table 4.20: Results of the Wilcoxen Rank Sum Test in Control and Experimental Groups 276
Table 4.21: Analysis of Pre-Test to Post-Test Responses in Control and Experimental Group in a Single Study (Rose Group) 278
Table 4.22: Odds Ratio (OR) with 95% Confidence Interval (CI) of BLTHQ Items 280
Table 4.23: Reliability of All Factors in the Control and Experimental Groups 285

Table 4.24: Principal Component Analysis (PCA) of Post-test Questionnaires of Control Group 286
Table 4.25: Principal Component Analysis (PCA) of Post-test Questionnaires of Experimental Group 286
Table 4.26: Student Weighted Averages Scores on Epistemic Stances and Consistency Scores of Control Group 289
Table 4.27: Student Weighted Average Scores on Epistemic Stance and Consistency Scores of the Experimental Group 291
Table 4.28: Student Consistency Scores on Epistemic Stances of Control Group and Experimental Group 296
Table 4.29: Categories of Low, Moderate and High Consistency Score of 299
Table 4.30: If Two People are arguing about something, at least one of them must be wrong 303
Table 4.31: Some People Just Have a Knack for Learning and Others Don't 304
Table 4.32: Percentage Difference of Items in the Control and Experimental Group 305

LIST OF FIGURES

Figure 1.1: The Theoretical Framework 8
Figure 3.1: Features of the Research Design..... 104
Figure 3.2: Map of Research..... 110
Figure 4.1: Comparison of Pre and Post Responses in Control Group 219
Figure 4.2: Comparison of Pre and Post Responses in Experimental Group 219
Figure 4.3: Comparison of Post Responses of Control and Experimental Group 220
Figure 4.4: Difference-in-Difference Estimation of Control and Experimental Group 225
Figure 4.5: Comparison of No Change, Degree Change and Category Change Stance of Control and Experimental Groups 229
Figure 4.6: Comparison of Low and High Stance Control Group 233
Figure 4.7: Comparison of Low and High Stance Experimental Group 234
Figure 4.8: Comparison of Change in Control and Experimental Group 268
Figure 4.9: Comparison of Pre-Test to Post-Test Change of Copier, Criterialist and Borrower Stance in Control Group 273
Figure 4.10: Comparison of Pre-Test to Post-Test Change of Copier, Criterialist and Borrower Stance in Experimental Group 275
Figure 4.11: Comparison of Consistency Score of Control and Experimental Group 299

CHAPTER 1

INTRODUCTION

'The problem provides the occasion and enduring focus of experimental inquiry.'

Scheffler (1999)

This introductory Chapter begins by outlining the problem of interest that provides the focus of this research in teacher education. Key ideas and perspectives that shaped my thinking and the development of the theoretical framework for this research are briefly discussed. The relationship of a perception of agency and change of concepts that is the focus of this research is introduced in the context of the evolution of ideas about conceptual change beginning with Kuhn's theory of paradigm change. The research focus is situated in theoretical frameworks on conceptual change, specifically, epistemic belief change and this is discussed, explaining decisions for the methodology of research. The argument is introduced contextually, elaborated and key issues are highlighted. The purpose, benefits, importance and scope of the research is explained, highlighting what could be seen as contribution to knowledge. A conclusion sums up important themes and outlines the direction that is taken in the methods used in the study.

1.1 Introduction to the Problem

Expectations of teacher education include helping teachers review and revise deeply held concepts and beliefs such as concepts about how people learn or the role of the classroom environment in affecting pupil behaviour. Valuable time is spent in in-service professional development of teachers to change and develop such key educational ideas. I, Lubna Kidwai, the practitioner researcher, teacher educator, have observed that the dynamics of whole group, teacher led discussion creates a seemingly shared social-construction that is not always a reliable

indicator of change in individual views or concepts. A year spent in interaction with a classroom group in discussions builds hope that change has indeed taken place but a visit to schools reveals traditional classrooms with desks in rows and beliefs in received knowledge as the dominant philosophy.

Perhaps, I wondered, if it was the discourse in our teacher education classrooms that was misleading. Perhaps, we the tutors, were reading more into the discourse than what lay underneath. Student teachers may be expressing a socially constructed stance that was merely surface learning and did not really reflect a deep belief change. Or, it could be, that the product of group learning did not reflect individual belief change. What is needed I recognized, is research that yields theory to better explain classroom learning of concepts in contexts of task based discourse in teacher education classrooms. Task based discourse is a term I have coined to refer to discussion that takes place in groups in the context of a learning task. My objective is to make these vital occasions more fruitful for change in concepts.

In order to understand the phenomenon, I turned to the literature for an analysis of the history and processes of thinking in conceptual change. An in-depth review highlighted interesting aspects of Kuhn's structure of paradigm change[4] that I could see were incompatible with contexts of learning as prevalent in classrooms. Apart from insight from theory, reflection on my experience of teaching raised questions that required answers. A chance observation of a small group of early years children occupied in learning seemed to illuminate constraints in the task based group discourse that may have hindered learning for some. These were children working in groups but I could see constraints and opportunities that might be found in teacher learning groups.[5]

This study problematises classroom interaction as unequal in its affordances of agency for individuals to reflect on prior beliefs and the new ideas on offer. A question is raised if concepts change more easily where there is an enhanced perception of agency in task based group discourse. The task based

[4] Thomas Kuhn (1962)
[5] See Appendix A for the Case Study.

discourse provides the necessary contexts, actions and words, but is reflection on own prior ideas possible within this scenario? I base my argument for agency on a review of Kuhn's analysis of paradigm change, and look for answers in Schön's work followed by Newman's, on ways to resolve the problems of convergence in meaning. Sinatra and Pintrich (2003) give importance to intentionality and motivation as a factor in learning and this too informs the development of the argument. The construct, a perception of agency and its relationship to conceptual change in contexts of task based group discourse is thus developed as a focus for my research.

1.2 Theoretical Background in Teacher Education

Teacher education as conducted in large private school systems in Pakistan includes formal courses of study in education conducted in collaboration with UK universities. As a teacher educator involved in such programmes, I chose to focus on a major issue of the development of teacher concepts. Conceptual change, while being an aim of teacher education, is often difficult to achieve.

Mayer-Smith and Mitchell (1997, p.129) find mixed results in promoting change in pre-service teachers' beliefs and practices. They discuss the failure of teacher education programmes to impact on the prior knowledge pre-service teacher candidates possess. They conclude that even robust efforts of those involved in carefully crafted programs aimed to promote conceptual change have minimal effect. In our own practice of teaching university based courses to teachers in Pakistan, we find that teaching and encouraging reflection has remained problematic. Zeichner and Liston (1987) discuss the goals of reflective teaching to argue for the importance of greater teacher autonomy and democratic participation.

Richardson (1997, p.6), discusses Schifter and Simon (1992), to suggest that the negotiation of meaning within social interaction often provides a source of cognitive dissonance that allows individual students to restructure their concepts. Loughran and Russell (1997, p.180), in describing their pre-service teacher

development programmes, say they strive to use a pedagogy that challenges new teachers to better articulate and understand their beliefs and practices which are rooted in prior experiences. This ‘better articulation and understanding of personal beliefs and practices’, I have come to realize, is key, especially in the contexts of social interaction in educational programmes. There is a need for more research in order to understand the phenomenon of the individual's agency in task based group discourse to articulate ideas and make connections with prior beliefs and concepts. Procedures that provide opportunity for more democratic and agentic discourse need to be understood and practiced to improve the possibilities of change.

1.3 Competing Explanations of Conceptual Change

Study of concepts and conceptual change, including individual beliefs about knowledge and knowing, is an area of interest in education and psychology, which is traced to the work of Piaget in genetic epistemology and Vygotsky on social constructivism. Change in concepts is studied with reference, among others, to the theories of Kuhn (1962) on paradigm change, Posner, Strike, Hewson & Gertzog (1982b) and a range of literature in cognitive psychology including the work of Sinatra and Pintrich (2003) and Vosniadou (2003). Change in epistemic beliefs is a focus of research in the cognitive sciences.

Change is seen as occurring in the individual mind in the Piagetian perspective, as assimilation or accommodation of new ideas with prior knowledge. The key finding of ‘How People Learn’ research is the importance of student’s preconceptions, say Ashby et al. (2005, p.79). This, I argue, is imperative. We teachers need to recognise that prior conceptions of each individual may be different and not necessarily common with and homogenous to those in the group. There may be varying levels of conflict and difficulties in accommodation. Social and cultural factors play a role in the process. Posner et al. (1982, p.212) argue learning as a rational activity. They refer to Kuhn (1970) and Lakatos (1970) to support patterns of change in concepts as comparable to descriptions of paradigm

change in science. Based on Kuhn's theory of paradigm change, conceptual change even at an individual level, is construed as arising out of conflict with existing beliefs.

There are two main approaches to the study of conceptual change. One is that of cognitive development and meaning construction, as in Piaget, the other of knowledge as socially constructed in discourse communities, O'Loughlin (1992), Hofer (2002, p.25). Vosniadou (2007a, p.1) describes a reframed, constructivist approach to conceptual change. Vosniadou (2007b) considers knowledge to be organized in domain specific, theory-like structures, and knowledge acquisition to be characterized by theory like changes as in Kuhn's theory of paradigm change.

Pintrich (2002, p.403), in a review of research, proposes that epistemological development is a function of both, internal psychological mechanisms as well as contextual facilitators and constraints. This is discussed in the review of literature and informs the development of the focus on a perception of agency in task based discourse. Change in concepts is considered a psychological process and contextual factors are seen to have meaningful effects. Pintrich et al. (1993, p.167) highlight a need to focus beyond cold cognition, on motivation and the roles of individuals in a classroom learning community. Change as intentional is also discussed with reference to the key voices in the domain.

Vosniadou gives importance to prior learning. While accepting the importance of socio-cultural factors, she suggests that 'radical socio-cultural or situative perspectives 'consider only the internalization or appropriation of existing cultural practices, tools and artifacts, and do not give adequate attention to the active role of the individual in understanding or constructing new knowledge'. Vosniadou (2007a, p.3). She believes that the conceptual approach could be reframed to account for both.

Such ideas challenge an uncritical paradigmatic belief in social constructive learning situations benefiting learning. There is faith that the individual who has to construct knowledge with reflection on prior concepts will best be able to do so in social interaction. I too share the faith but have learnt in the years of my practice that there are factors at play in group work and interaction that need to be explored

with more research in order to improve the possibility of change in concepts, each individual at a time. Scaffolding learning in task-based discourse may require more awareness of difference in individual prior concepts and efforts to encourage individual agency to reflect upon these within the social contexts of learning.

It is useful here to note developments in research and practice in the 'Learning Sciences', which is a move away from 'Instructionism' (Sawyer, 2006, p. 1-16). Importance is given to deeper understanding of complex concepts and creating effective learning environments. Importantly, these are argued as constructivist approaches that emphasise prior knowledge, scaffolding, externalization, articulation and promote reflection for better learning.[6]

1.4 The Theoretical Framework: An Argument for Agency

There is a large and diverse body of research and theory that informs the work in conceptual change in learning. Kuhn's theory of paradigm change is influential in building theory in the classical approach to conceptual change. Posner et al. (1982); Sinatra and Pintrich (2003, p.7); Ferrari and Elik (2003, p.33); Vosniadou (2007b); Thagard and Zhu (2003, p.79); are among those who argue change from a Kuhnian perspective.

While Kuhn's theory is often cited in beginning discussions on conceptual change, in my understanding, Kuhn's theory of change in paradigms could be incommensurate with change through instruction. Instruction in classrooms has a change agent and planned change is assumed first and then brought into production. Therefore, to import Kuhn's theory to explain and predict conceptual change in classrooms one must recognize and accommodate the difference. The autonomous, cognitive/socio cultural processes that make existing ideas obsolete and generate consensus and enthusiasm for the new, creating a new paradigm, are not part of planned classroom process. Change through instruction is

[6] See
1. Sawyer (2006, p.1-16) in The Cambridge Handbook of the Learning Sciences (2006)
2. How People Learn; Brain, Mind, Experience and School (2000)

generated deliberately by others in artificial situations and on possibly unsuspecting actors. Classrooms are places for more 'mediated action' as opposed to the autonomous, naturally evolving action that is taken as examples of paradigm change.

It is in this context of complex learning that Kuhn's analysis of changing paradigms can be revisited to account for change even at small scale and individual level. The change in paradigms as social-cultural events, have been, as I have highlighted earlier, autonomous, often hotly debated, emotionally charged. Issues are real and the change is often radical and especially in the domain of science, the change has been effective over large groups and over time. In classrooms, issues may be of transient, cursory interest. Participants may not see the need to invest energy in what they may see as an interest of the group rather than a personal concern.

Machamer (2007, p.39) discusses Kuhn's philosophical claims to agree that Kuhn was right to tie knowing with acting and doing. He argues that knowledge use is in the public domain which entails that criteria for knowledge always has a social component. Machamer (2007, p.39) refers to Piaget's accommodation and assimilation to argue that knowing and doing involve recall of prior schemata with feedback received in the environment leading to 'knowing well'. This, as a long practicing teacher, I can argue, is where the problem arises. Knowing is simultaneously an individual, psychological action, with the individual expected to make connections with prior schemata but in social situations which are necessary yet at times possibly obstructive, in allowing this process to take place. My argument is that a perception of agency may be present or missing in the social situation which may impact conceptual change. The video record of an early years' classroom that is provided, illustrates just such a situation.[7]

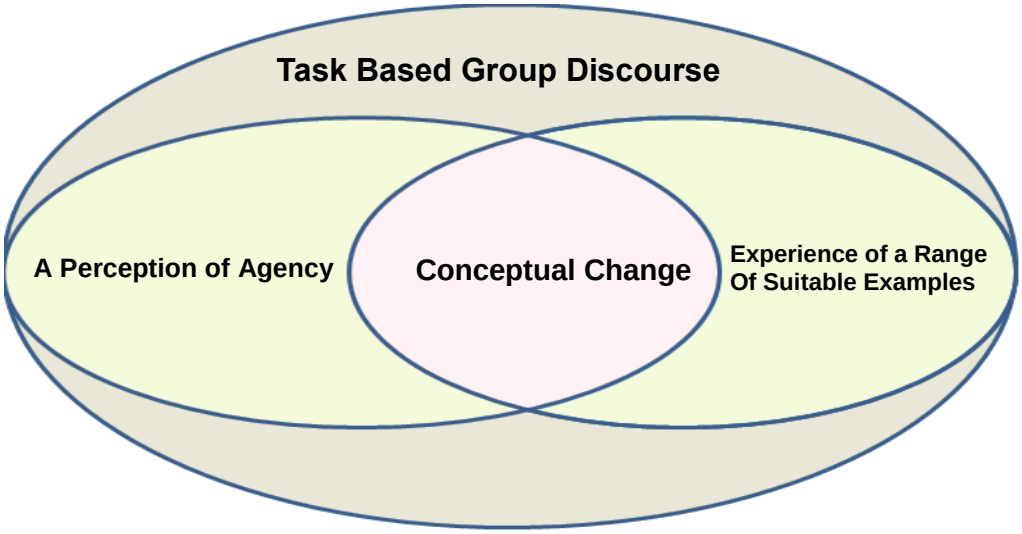
[7] See Appendix A1.

1.5 Pedagogical Method for the Experiments

In order to research a perception of agency in practical classroom teaching experiments it is necessary that all other variables are controlled. The pedagogical method needs to be common in all experiments and clearly defined and described. The method also needs to be effective for learning in order for the teacher to be able to focus on the environment of a perception of agency to change concepts. For this research, Skemp (1971) principles of change in concepts in the domain of mathematics were selected. This method is described and is itself examined for effectiveness in change of epistemic beliefs in history.

The theoretical framework developed on the basis of these arguments is illustrated below in a Venn diagram set in a context (See Figure 1.1). Conceptual change is hypothesized as emerging from an experience of a range of suitable examples supported with a perception of agency to question, reflect on own prior knowledge, agree or disagree with group views and freedom to circulate to a more compatible group. The context is task-based group discourse in teacher education classrooms.

Figure 1.1: The Theoretical Framework



1.6 Research Aim

To enquire into a problem of teacher learning as change of concepts that I and my colleagues encounter as practitioners in teacher education.

1.7 Study Question

The main research question is, if there is a relationship between a perception of agency and change in participants' epistemic beliefs in learning contexts of task based discourse.

1.8 The Research Hypothesis

The hypothesis states that there is a relationship between a perception of agency and change in concepts in task based group discourse.

1.9 Key Objectives of Research

- To investigate an alternative hypothesis of a relationship between a perception of agency and conceptual change in task-based discourse in 5 quasi-experimental studies conducted in teacher education contexts.
- To employ and examine Skemp (1971) principle of a range of suitable examples as a pedagogical method to change concepts.
- To demonstrate if the relationship of a perception of agency and conceptual change in task-based group discourse can be researched.

1.10 Purpose of Study

This research is planned and conducted in partial fulfillment of the requirements of research for the award of a Doctor of Philosophy from the University of Bradford. As I am an extra mural candidate, this independent, self-funded, research has been conducted in my own country, Pakistan, in the field of private schools providing teacher education facilities. Distance learning has its challenges but the constant support from the University and of academics in the field as well as provision of resources, makes the task possible to accomplish as this study demonstrates.

The purpose of the study is to enquire into a problem of teacher learning and change that I and my colleagues encounter as practitioners in teacher education. Most teachers in Pakistan come from a background of traditional schooling. Traditional schools have teacher led, whole class teaching, one correct answer to questions that are rote learnt and reproduced in examination. Behaviour and learning is conceived as a consequence of upbringing and societal influence. Concepts that teachers bring to the teaching encounter are notoriously hard to change and develop, as I and my colleagues have found. This background of concerns about practice has informed the development of my hypothesis and research project, and is of personal interest to me. Conceptual change is a growing field of study in educational psychology and a new relationship with change may be of general interest in the community of teacher education as well as in other domains. A personal, intentional interest to research my practice and its problems of conceptual change was the initial impetus for the choice of a research degree.

Practitioner research has its common practices including action research, case studies and qualitative methods. This research conducted in my practice of teacher education demonstrates the use of quasi-experiments and mixed methods. As a teacher educator, one of my objectives is to model other ways of conducting practitioner research for my teacher students in Pakistan.

1.11 Benefits of the Study

Research in the field of conceptual change should benefit from the proposal of a new relationship for study. Change is researched from psychological perspectives of cognitive development, from a focus on cognitive dissonance as creating conditions for change, as well as from social and situational perspectives and there is a range of literature[8] that inform this field. The review of literature, especially that of paradigm change in scientific communities, introduces a critical analysis of the process of paradigm change as interpreted by theorists in the field of conceptual change following Kuhn. As an outcome of this analysis, I introduce a perception of agency as a possible factor in change and my research looks for support for this relationship. An alternative hypothesis is proposed and found to be supported in the research. This could be followed by other research to reject a null hypothesis.

Conceptual change is considered the most central area in the learning sciences as diSessa (2006, p.265) points out. Supporting change through task based group discourse in classrooms in a range of concepts in domains such as science, history, mathematics, and geography can be explored with a perception of agency as a factor in change. A relationship of change and a perception of agency can also be studied in contexts of organizational change.

Epistemic beliefs about knowledge in history are an important area of research. Measures are still in the process of being developed. The CBKH, (Categories of Belief about Knowledge in History) rubric for assessment of epistemic beliefs about knowledge in history was modified and developed in this study and employed in the assessment of 166 responses. The rubric is trialed and found useful and can benefit from development in more research in the field.[9]

[8] See Sinatra and Pintrich (eds) (2003) Intentional Conceptual Change; Sawyer K. (ed.) (2006) The Cambridge Handbook of the Learning Sciences; Donovan and Bransford (eds) (2005) How Students Learn History Mathematics and Science in the Classroom.

[9] The rubric was developed with reference to Kuhn and Weinstock (2002) Levels and Maggioni et al model of the Copier, Borrower, Criterialist categories of belief.

The pedagogical method of using a range of suitable examples finds strong support with significant conceptual change observed in 10 groups. This is adapted from Skemp (1971, p.25, 32) to experience a collection of suitable examples as necessary to learn mathematical concepts. Further trials in primary school learning of science, social studies and mathematics concepts could draw attention to this useful pedagogical tool. I have found that the three specific conditions of experience, range, and suitability of examples are all necessary for making a difference in change. This is another benefit of the study pointing towards more areas of research.

The word collection may be taken to imply a group of collected objects or examples. Range implies the amount, number or type between an upper or lower limit. Skemp (1971) discusses examples and non-examples of concepts which is a useful distinction. I provided types of sources of evidence in the workshops and these were discussed with reference to questions. Details are provided in the Chapter on Research Methodology.

One interesting aspect of this research is that the particular concepts identified for study are epistemic beliefs about knowledge in history. These concepts and change within these are the source of requirements of teacher development in line with change in Syllabus of the Cambridge O' Level Examination (CIE) of History in Pakistan. I have identified this requirement of teacher education from a reference to targets of teacher development in the UK and elsewhere, these not being stated as such in academic manuals of the private schools, where Cambridge examinations are conducted in Pakistan, as yet, to my knowledge.

History teaching and learning is a politically dictated activity in most parts of the world although curriculum improvements show a trend towards development.[10] History teaching content in Pakistan usually follows changing patterns with each changing government and provincial body. Local examinations require a limited syllabus definition and the scope of development is not large. Cambridge

[10] A quick Google search, with Pakistan Studies or Pakistan History as search terms, will reveal the intense interest in the subject on an international level as well as within Pakistan.

International Examinations are set to international standards, therefore change and development can be expected for the better. My research should be useful to inform CIE that teachers can develop epistemic concepts and that change of the syllabus can lead the way. To play a positive role in developing critical thinking is an important aim of education.

Quasi experiments may appear to be somewhat technical and bound by theory in practice; however, all teachers try out new ideas and methods in their practice and indeed many have their bag of tricks developed in much the same experimental ways. Interestingly, mixed methods for all their paradigm challenging theory are, again, something teachers and tutors use all the time whilst assessing and reporting on pupil achievement. It would benefit teachers to learn the *limitations of inference* from less than reliable methods and different types of methods, measures, and analysis, to draw useful conclusions. The problem of conceptual change that is identified is one I and my colleagues in teacher education are familiar with, as indeed would be teachers in schools. My own role as a teacher practitioner researching my own practice allows me access to unravel in fine grain the many layered complexities that make up educational situations. Reflexivity allows the kind of insight that supports a reasoned inference that is necessary in quasi experimental research.

The development of the rubric, the CBKH, Categories of Beliefs about Knowledge in history may be of benefit to future researchers interested in the critical analysis and development of such instruments. This is further elaborated in the Section that follows.

What I learn from the conduct of this research will support reflection on my own learning, as indeed it has immensely, as well as I hope, be of interest to inform the domain of teacher education generally. Conceptual change is a complex phenomenon that has much potential for study as discussed in the review of literature. What I have learnt about epistemic beliefs in the years of this research is indeed a life changing experience and I would like, Quixote like, to run out in the field and warn people rather than just sit back and assess their claims to knowing.

Truth is a fascinating subject and the years of this study have been a most personally rewarding learning experience.

1.12 Importance and Scope of the Study

A key contribution of this study is the proposal of an interesting relationship of a perception of agency and conceptual change in task-based group discourse. The argument is novel, emerging as it does from Kuhn's description of the change of paradigms and concepts within these. Kuhn's theory of paradigm change has long been considered the basis for the development of theories of conceptual change through educational experiences. Differences in a perception of agency are highlighted in the *context* of change in scientific communities and in task based group discourse in teacher education classrooms. This is an important factor that needs recognition in planning for change through instruction in both child and adult education as my study suggests.

Conceptual change is a major concern in the field of education and in its various domains. Emerging from Kuhn's claim of incommensurability between paradigms, much theoretical argument and research tries to address the problem of convergence in meaning. Schön's work to resolve the problem of meaning by emphasizing contexts, actions, words and reflection in action is influential while being disputed by others such as Newman (1999). A perception of agency as making a difference within the context of task based discourse fits in this space.

Change in concepts is studied in Science, e.g. Driver (1983, p.180), in Mathematics, Skemp (1971), History, Lee et al. (1993), Shemilt (1980), VanSledright (2002), Lee (2005), Maggioni et al. (2009a), VanSledright (2014) and in a range of other fields and publications by the authors.[11] Measures of epistemic beliefs are several as Duell and Schommer-Aikins (2001) analyse.

[11] See Donavan and Bransford (eds) (2005) How Students Learn; Myint Swe Khine (ed) (2008) Knowing Knowledge and Beliefs; Hofer and Pintrich (eds) (2002) Personal Epistemology; Vosniadou, Baltas and Vamvakoussi (eds) (2007) Re-framing the Conceptual Change Approach in Learning and Instruction.

My arguments fit in this theoretical domain and introduce the factor of a perception of agency in the social context of task based discourse as playing a role in the process of an individual's surfacing of own prior concepts, recognising conflict and finding new ideas acceptable or useful. This should be a fruitful new avenue of research in conceptual change. The advantage is that it is logical and fits in well with arguments such as those for other psychological constructs including intention and motivation as relating to change. The theoretical background of agency and mental development of infants and young children goes back to Piaget (1928) and Russell (1996). It may be worth exploring if a *perception* of agency is a factor in adult learning of new concepts. Qualitative research could shed more insight on this relationship.

There is scope in this argument for research into change of all kinds of concepts not just epistemic beliefs about knowledge in history. Epistemic beliefs and change within these are studied in a range of domains by researchers around the world. Conceptual change is of interest generally in management of change in organizational cultures as well as cultural change in societies. Change in concepts in society is produced intentionally as well as unintentionally through a range of processes and media and for a wide variety of purposes. Perceptions of agency in processes of change especially where this is orchestrated by agents on unsuspecting subjects may be usefully taken into account. It would be interesting to see if a perception of agency could be a common denominator in societal change as well as change in the contexts of interaction in classrooms.

Somekh (2006) discusses agency and change to generate actionable knowledge and action research for agency in organizational change. Somekh considers the constraints in institutional structures and the affordances that either enable or constrain the action of individuals to bring change. In this research I focus on a perception of agency that bridges a psychological accommodation of new ideas to change existing schema within the context of task based discourse. My focus is on a change of concepts and not on action or bringing about change in organisations, however I see scope for research within organisational change

frameworks of a perception of agency bringing about measurable difference by simply helping to change the way people think.

Demonstrating that the relationship of agency with change can be researched in practical terms is particularly useful. This is practitioner research using a quasi-experimental methodology. The trial is not a randomized control trial; however, it has its value in validity for similar contexts and samples. Conducting 5 studies with 10 Groups, Control and Experimental, based on this relationship provides reasonable evidence that the study can be replicated by other practitioners in classrooms. Operationalising the perception of agency is a useful model for future use.

The design of the research is reported with transparent detail and should be useful for teacher educators and researchers to follow. Conducting 5 matched quasi experiments and carrying out independent as well as meta-analysis of all 5 studies was a useful if time consuming methodology and this is described in detail. This would have benefits for others who carry out research in education, where there are debates about how valid knowledge can be generated by practitioners. The constraints and opportunities of the context described forewarn researchers about the limitations of research in the field as well as its value for practice.

Epistemic beliefs of teachers about knowledge in history are explored through a qualitative analysis of responses to open ended questions. These responses are assessed with reference to a scale of epistemic beliefs developed for this purpose. Kuhn and Weinstock (2002) as well as Maggioni et al. (2004). Maggioni et al. (2004) models of epistemic beliefs (2004) are used to build a revised set of categories to order responses. This rubric called the CBKH or Categories of Beliefs about Knowledge in history is trialed, amended and improved with repeated readings of the 166 responses sheets. Exemplars are added to illustrate the categories. This instrument might be seen as useful by others researching epistemic beliefs about knowledge in history. As this is a new development, there is scope for critical analysis and more research.

This research finds that concepts are learnt well with experience of a range of suitable examples. In the 5 studies, all 10 groups showed significant change

from pre-test to post-test. This is another good theory to examine in conceptual change research. Skemp's theory for the use of a collection of suitable examples for change in concepts is well argued, Skemp (1971, p.32). His idea always made sense to me and employing the method formally for research with careful planning and implementation proved effective. Making sure all three examples were used in the workshops was not hard. The qualifying adjective 'suitable' turned out to be key, as the range of examples I used were meaningful, had good explanatory value, and caught the imagination of the participants. Therefore, the words, experience of a range of suitable examples are all of importance. Experience is also an important element here. The teacher participants were provided concrete experience of the source materials, examined these carefully and drew conclusions that were meaningful to them. Significant change, as seen in both Groups, Control and Experimental, bears testimony that the method did work. The examples and heuristics used are employed as argued by VanSledright (2002); VanSledright and Afflerbach (2005); Lee and Shemilt (2003).

Teaching to change concepts is intentional, where teachers are clearly aware of the scope of the goals they plan. How such teaching can be made effective, whilst still humane, moral, not manipulative and not coercive, is a complex issue. Agency in intellectual activity is not counter intuitive. This research suggests that there could possibly *be* a relationship between a perception of agency and change in concepts in task based group discourse. Purely logically, it would have been strange indeed if the inquiry had suggested otherwise.

Finally, this quasi experimental research using mixed methods, is conducted taking a generally pragmatic stance. I question 'what works', however and suggest that 'what is known to work' is of value only if it is known how it is known and to what parameters it can be held to apply. The research is conducted as 5 independent studies and detailed transparent and open reporting is provided for other researchers in the field to try out.

CHAPTER 2

THE REVIEW OF LITERATURE

‘Even whatever is written in the book may be totally wrong like aa Bahadur Shah Zafar was the last emperor and...’

‘It’s true, that is the thing I am telling.’

(Loud buzz from group)

‘Ma’am was asking that aa are you sure about that?’ *(giggles, laughter)* ‘we said yes we are sure about that. And she is asking, how do you know and we said....’(participants joining in)

‘it is written in the book.’ (laughter)

- Participants in a focus group session reflecting upon change in their epistemic concepts in the history workshops. The ‘Ma’am’ they refer to is myself, the tutor.

2.1 Introduction

This Chapter is composed of essays that critically review relevant literature that has led to the development of the theoretical framework, informed my understanding of philosophical and practical aspects of the design, issues in conceptual change particularly epistemic beliefs, and importantly, the development of the CBKH rubric, the Categories of Belief about Knowledge in History. The development of the research design required much reading and reflection on issues of paradigms in mixed methods, reliability in quasi experimental research, and indeed the drawing of conclusions from the mixed data. All of these are important reading and have shaped my understanding as well as helped me identify gaps in knowledge and thought that research efforts could serve to illuminate. I would like to draw special attention to Sections 2.2.6 to 2.2.8 that argue and present the key relationship of agency with change in concepts in the context of Kuhn’s ideas of how paradigms change.

The focus being complex and the literature vast, it requires far more coverage than is possible in this literature review and word limit. I have been selective and focused on key issues and arguments relevant to my research. I explore knowledge in these areas and proceed to critically select, review and add my voice to the debate. The first section traces the development of understanding of learning as conceptual change with its basis in Piaget's genetic epistemology, the history of science approach, and Kuhnian paradigm change. Conceptual change, specifically epistemic concepts and change within these, concepts of knowledge in history, agency as a factor in change and the history of research within this domain, is critically reviewed. The relationship within these is explored to find relevance and meaning and indeed to gain insight into the complex phenomenon of conceptual change in teacher learning.

The research focus, a relationship between conceptual change and a perception of agency in task-based group discourse is at a nexus of knowledge in domains of cognitive psychology, philosophy, sociology and education. That sounds more grand than it is. In everyday classroom practice in schools around the world all of these areas interact and are the focus of concern of learners, teachers, schools, parents and the society at large.

2.2 Concepts, Epistemic Beliefs?

The focus of interest in this research are concepts, specifically, epistemic beliefs about knowledge in history, and change within these. In this section of the literature review and those that follow, what are epistemic beliefs, how change in such concepts can take place, and aspects of the social context that may make a difference in change are explored in depth. I raise a key weakness in existing understanding about the relevance of Kuhn's description of paradigm change and the process of change of concepts in classroom situations to identify a possibility that a perception of agency may be an underpinning difference in the change processes that take place in scientific communities and the orchestrated change in concepts in classrooms.

diSessa (2002, p.32-33, 58) is critical of research that is built on vaguely defined terms such as ‘concept’, ‘belief’, ‘theory’ and ‘ontology’. diSessa argues in picturesque ways, that the lenses to inspect conceptual change are ‘fuzzy’ and what we have in theoretical terms is a ‘mushy soup’. In place of coherence and broad terms such as concepts, she takes the position that concepts be seen as fragments, as parts of coordination classes and recommends a “conceptual ecology approach”, which is a complex organisation or reorganisation of different kinds of knowledge. I discuss this further below.

Hewson (1992, p.8) sees epistemological beliefs as part of a person’s “conceptual ecology”. Chi and Roscoe (2002, p.6-7), suggest that one can represent knowledge as a set of interconnected ideas as a ‘mental model’. This adds a structure in which these propositions are embedded. They say there can be fragmented or coherent models. Schema, mental model, theory, theory-like, are more terms employed in the literature.

Pintrich (2002, p.390) discusses the nature of the construct of personal epistemology expressing concern that diverse models discussed in the literature ‘might not be concerned with the same construct or at a minimum might not define the boundaries of the construct in the same way’. A review of literature finds that the terms 'personal epistemology or epistemic beliefs' are complex with a broad range of contextual and domain related meanings. Hofer (2002, p.4) frames epistemology from a philosophical perspective, as being concerned with the origin, nature, limits, methods, and justification of human knowledge. This I consider a particularly useful definition in the context of knowledge in history. Hofer (2002, p.4) differentiates the term ‘epistemic’ as relating to knowledge more generally, and to the conditions for acquiring it. Kuhn and Weinstock (2002, p.123) argue that well developed epistemological understanding would demonstrate a coordination or balance between the subjective and objective dimensions of knowing. This is good insight as I have seen responses demonstrate development with a growing awareness of the constructed nature of knowledge followed by understanding that the use of heuristics can help find the better explanation in history.

I explore what are epistemological beliefs about knowledge in history, and a ‘perception of agency’ in this Chapter in depth. In the description of the research I will spell out the frame of reference for ‘a perception of agency’.

2.2.1 Perspectives on Conceptual Change

Learning as change of concepts is a concern in education. A Report on ‘How Students Learn’, in texts produced by the National Research Council of National Academies in America, begins with the statement that students come to the classroom with preconceptions about how the world works and, if their initial understanding is not engaged, they may fail to grasp the new concepts and information, or they may learn them for the purpose of a test but revert to their preconceptions outside the classroom. Donovan and Bransford (2005b, p.20) raise the concern that revealing prior ideas and changing these in the course of instruction is essential to effective learning and receptive teaching. They find that a focus on student thinking would as a norm encourage expressive behaviour as well as risk taking. Donovan and Bransford (2005a) also suggest that mistakes should not be judged as evidence of poor learning but rather as demonstrating understanding. I draw in part on these ideas to operationalize the construct of a perception of agency as an intervention.

The effectiveness of teacher education programmes in changing teacher beliefs and expectations such as that all children are capable of learning and that teachers hold and communicate their expectations regarding children's learning are questioned by McDiarmid (1993, p.113). In my own experience, I know that while teachers enjoy listening to accounts of Rosenthal's research, the thought that their *own* expectations could be making a difference seldom sinks in. A four year longitudinal study in eleven programs found that teachers’ ideas about ethnic stereotypes being valid knowledge on which to make classroom teaching decisions did not change in most of the sample. However, they did find statistically significant differences on a number of items, McDiarmid (1993, p.137).

These are daunting statistics but research must go on if concepts are required to change. In this research as well, I find that while there is good change of a degree or from low to higher stance, change of a complete stance is infrequent in the sample.

A view of conceptual change as a process of equilibration through accommodation owes its clarity to Piaget. There are two main *approaches* to the study of conceptual change. One is that of cognitive development and meaning construction as in Piaget, the other of knowledge as socially constructed in discourse communities. O'Loughlin (1992), Hofer (2002) and Scott et al. (1997) review strategies to promote conceptual change. They identify two approaches; cognitive conflict-resolution of conflicting perspectives; building and extending learners existing ideas. Roth sees an integration of psychology and anthropology in Vygotsky's approach. He terms this a socio-cultural and socio-historic approach which integrates both cultural and individualistic ways of knowing. Roth (1999, p.13).

Pintrich (2002, p.403), in a review of research, finds that both internal psychological mechanisms as well as contextual facilitators and constraints have a role in epistemological development. According to Pintrich (2002, p.403) 'strong cognitive developmental claims about internal cognitive processes (e.g., equilibration) or strong contextualist claims about socialization processes (e.g., scaffolding) as the sole or main mechanisms of change are not warranted.' Pintrich (2002, p.404) concludes that there is no need for research that compares internal psychological mechanisms with contextual factors as facilitators of epistemological development.

From the perspective of psychology diSessa (2006, p.266) argues that 'there are, in fact, no widely accepted, well-articulated, and tested theories of conceptual change'. diSessa (2006), analyses what she calls threads and fault lines in the history of theorizing and research in conceptual change. Concepts are seen according to diSessa (2006), as coherent and strongly integrated or as fragmented, quasi-independent elements. Vosniadou (2007a, p.1) considers knowledge to be organized in domain specific, theory-like structures, and

knowledge acquisition to be characterized by theory like changes as in Kuhn's theory of paradigm change.

diSessa, argues that the term "concept" is 'hopelessly vague' and urges more precise definition. She promotes a move towards a study of conceptual change in finer grain size, multiple and, according to her, therefore, more accountable concepts. She recommends studying concepts as a greater number of small scale, diverse elements. In place of the broad term concept, diSessa advocates the use of 'phenomenological primitives' or 'p-prims' organised in 'coordination classes' but recognises that if concepts are seen to exist in the thousands within conceptual ecologies there would be an issue to know which are 'activated' in particular contexts. diSessa considers that context dependence should be expected, diSessa (2002, p.29-51).

In my initial understanding, if a researcher were to employ diSessa's idea, what would necessarily be needed in a review of research is to find sound empirical verification for such constructs. One would need to look for existing verified models that make specific links between each mental entity and the processes that are said to promote its development. As a reader of theory doing a literature review, I saw practical problems in researching smaller grain. I reasoned that given the range of possible cognitive/affective and socio-cultural factors and mechanisms in change, smaller grain in concepts may just as easily make causal relations difficult to establish and alternative explanations difficult to counter.

As my analysis of data progressed however, I learnt the usefulness of finer grain. I realized that indicators of belief required expansion, and elements that compose concepts of epistemological beliefs could be usefully unraveled for analysis. In the statistical analysis, exploring the change in concepts in finer grain as change between low and high stance, and no change, degree change, category change as well as a fine grained analysis of the neutral, revealed the underlying nature of the change as well as significant differences. This is described further in the Results and Analysis Sections.

Vosniadou (2007a) reinforces the view that conceptual change should not be seen only as individual, cognitive process but that the role of socio-cultural

factors be considered as well. Placing learners in circumstances where they experience cognitive conflict is seen as insufficient for change by Sinatra and Pintrich (2003, p.3). Motivation, affective resistance, and learner's beliefs are characteristics that they consider necessary while others take a socio-cultural approach or a balanced approach. Vosniadou (2007a, p.1) compares theory that conceives of learning as adding to knowledge where concepts are seen as enriched rather than changed with the constructivist approach that considers knowledge to be organized in theory like structures and knowledge acquisition to be characterized by theory-like changes.

While accepting the importance of socio-cultural factors, Vosniadou (2007a, p.3) reinforces the criticism of Hatano (1994), that radical socio-cultural or situative perspectives, e.g. Lave (1996), Rogoff (1998), 'consider only the internalization or appropriation of existing cultural practices, tools, and artifacts, and do not pay adequate attention to the active role of the individual in understanding or constructing new knowledge'. They believe that the conceptual change approach could be reframed to account for both. This stance is in keeping with the perspective I take in my study when I examine a perception of agency to reflect within the social context of task based discourse.

Ferrari and Elik (2003, p.24-25), discuss what concepts are from a range of perspectives. They identify meanings from standard experimental psychology, probabilistic theories, theory based models of concepts, relational theories and actional/situated perspectives. From actional/situated perspectives, they find that concepts are not localised in individual minds. Concepts in these perspectives, according to Ferrari and Elik, are abstractions that apply to people acting in social settings. I argue an either or approach as inadequate.

Vosniadou (2007) supports Baltas (2007) to argue that understanding a new concept should not require the replacement of a 'correct' theory with one that is 'incorrect' but 'rather the ability to move on to a new, wider, broader perspective'. This is in keeping with arguments in the history of science approach that see the new paradigm as a wider perspective informed by knowledge of the old. Vosniadou (2007, p.11) argues 'change should be seen as requiring the ability to

take multiple perspectives, examine different points of view and understand how they relate to different contexts of applicability'. There is, however, a concern in my understanding for instances where there is conflict between prior ideas and those being introduced. From a stance, for example, that history is isomorphic with the past to one that acknowledges history as a construction, beliefs are distinct and change from one set of beliefs to another is not easily accomplished.

Vosniadou (2007a, p.7) argues that the use of bottom–up additive mechanisms may not be useful in producing change through instruction. She concludes that conceptual change through instruction requires restructuring of children's naive theories, their models of learning, and the creation of meta-conceptual awareness and intentionality Vosniadou (2003). She concludes that in order to avoid misconceptions the use of appropriate tools, artifacts and situational contexts can be facilitated.

From the background of Kuhn's arguments about paradigm change, Posner et al. (1982) propose conditions necessary for conceptual change: there must be dissatisfaction with existing conceptions; a new conception must be intelligible; the new conception must appear plausible; and should be fruitful. Posner et al.'s model is widely used in conceptual change theorizing; Sinatra and Pintrich (2003) are, however, critical of 'cold conceptual change' models.

Bendixen (2002, p.191) confirms the role of epistemic doubt as a factor in change and considers that doubt and the struggle to resolve epistemic doubt is not an easy experience. Bendixen queries the role of the instructor in the process and recommends more educational research to better understand how to deal with the 'tumultuous experience'. In this research I examine the role of the teacher in developing and maintaining a perception of agency within task based discourse finding that to be a possible factor in change.

2.2.2 Conceptual Change in Various Domains

Apart from the subject of history, epistemic beliefs are studied in domains of science, mathematics, and in general, as well as at different age and educational

levels. Stages of development and relationships with culture, age, gender, education among other aspects are of interest to researchers.[12] Kagan (1992, p.73) discusses a sample of studies on teacher beliefs to conclude that teacher beliefs usually reflect the actual nature of instruction that teachers provide to students. My case study, provided in Appendix A-1, illustrates such an event.

Whilst framing conceptual change as domain general, Vosniadou suggests that conceptual change approaches make a larger claim about learning that transcend many domains Vosniadou (2007a, p.1). Domain specificity and generality of epistemic beliefs is debated in the literature with several studies located in particular disciplines such as mathematics, science and history. In one example, Clinchy considers that students may 'move from subjectivist to procedural knowing in the humanities but move from received knowledge directly to procedural knowing in science' Clinchy (2002, p.85). From the perspective of learning in science, Viennot (1979) and Driver and Easley (1978) support the ideas of Posner et al. (1982, p.212) of students changing their prior frameworks and replacing these with scientific concepts *in which they receive instruction at school* (italics mine) as related to the theory change approach of Kuhn. I see problems with this as I argue later in this Chapter.

Limon characterizes concepts in history as being different from those in science and history. In Limon's words, history concepts are abstract, implicit in narrative, have to be inferred; they change over time; consensus is difficult; categories are ill-defined; and it is difficult to construct reliable or global representations of concepts. According to Limon, there are important epistemological differences and science constructs are not applicable to conceptual change in history, Limon (2002, p.277-285).

For the purpose of my research I take the various points of view to draw a workable frame of reference as well as to critically analyse my findings. I am researching change in concepts of knowledge in history and Limon's argument has weight in this. In my experience in this research, a combination of narrative responses to open ended questions, along with Likert Scale questionnaires, is

[12] See De Corte et al, Bell and Linn, Elder, Qian and Pan, Hofer and Pintrich, 2002

useful to study epistemic beliefs in a population. Inferring a stance from a narrative over large numbers of responses is not an easy task but the use of a framework such as the one devised for this study, the CBKH, was most helpful.

2.2.3 The Use of Kuhn's Theory (1970) and Posner et al's Model (1982) in Conceptual Change Theory Building

Kuhn explored the history of science over periods of change to compare what he termed normal science and extraordinary science. According to him change in paradigms came about as a process of revolution rather than as an incremental, additive growth. Kuhn argued that science functions, as a rule governed practice with its peculiar concepts, methods and language in particular paradigms. Normal science is thus conceived according to Kuhn as a puzzle solving activity with additive growth. When anomalies are found that cannot be explained, crisis is said to ensue and the process of resolution of the crisis brings about change in ways of thinking and doing science. Puzzlement, awareness of anomaly, crisis, perception of novelty, resistance, large-scale paradigm destruction, major shifts in the problems and techniques of normal science, incommensurability, are states of affairs argued by Kuhn in the process of change in paradigms. As a potent example, Kuhn unravels in depth, the process of the discovery and recognition of oxygen in a comparison of Priestly and Lavosier's efforts to solve the puzzle where Lavosier was able to overcome the constraints of the existing paradigm to think in new ways whilst Priestly resisted change till the end of his life Kuhn (1962).

Kuhn's work has had a large impact on the development of theory in conceptual change. Ferrari and Elik (2003, p.33) consider that Kuhn's analysis of the history of science has been the inspiration behind the standard model of conceptual change in cognitive science. Relating Kuhn's theory of paradigm change to conceptual change work, diSessa (2006, p.268-271), cites a range of theorists including Karmiloff–Smith (1988), Nersessian (1992), Vosniadou (2002) in applying the history of science approach, seeing it's importance in the context

or process of conceptual change. diSessa (2006, p.266) discusses the ideas of Kuhn as seeing concepts as coherent and Toulmin arguing for fragmentation but supports the fragmented view as allowing disassembling, refining and reassembling. While I find a stance on fragmentation difficult, in my analysis I did find fine grain as diSessa argues but coherence overall.

Hoddeson (2007, p.26), Vosniadou (2007a, p.2) describe Kuhn in similar terms as seeing normal science functioning in a set of shared assumptions, beliefs, commitments and practices which he termed paradigms and change taking place as a theory replacement, a gestalt shift making old ideas incommensurable with the new. Science educators have found this framework fruitful to conceptualize learning of science concepts.

There are arguments in other fields about the extended use of Kuhn's structure where he may not have intended it. Kuhn's framework is socio-cultural, and diSessa (2006) questions its use to explain change at an individual level. London (1996) cites Capra to argue that paradigms imply change at the level of society and not at the level of the individual. London uses the term 'nevertheless' to say that the concept provides a useful metaphor for understanding the nature of change. I have found that while the rhetoric does demonstrate awareness of this contradiction yet a more constructivist approach to applying Kuhn's ideas according to fitness and purpose prevails.

Hoddeson (2007) discusses criticism of Kuhn to argue 'at the level of theory, there really are revolutionary changes and incommensurable differences between earlier ways of making sense and the ways that eventually replace them..' Hoddeson (2007) finds that Kuhn's theory of revolutionary paradigm shift has been used to describe changes lesser than those he analysed. She says however, that it has proved to be a useful model for researchers in the fields such as psychology and education. She discusses the influence of Kuhn in change towards thinking about science as a social activity that occurs in a particular period, expanding it to include other factors, Hoddeson (2007, p.26).

Cognitive psychology approaches are criticized by proponents of socio-cultural approaches. Vosniadou (2007a, p.2) sees the socio-cultural approach to

be actually in line with Kuhn. Vosniadou refers to Machamer (2007, p.3) that the 'introduction of the notion of paradigm, shifts the emphasis from individuals' minds to the role that the scientific community and their group commitments, shared examples, and tacit knowledge play in scientific discovery and change'. I find that while Kuhn expanded the notion of science from beliefs and practices of individual scientists to those of a community, he did consistently provide examples of change at the level of individual minds. He expounded the notion of meanings themselves being internal to paradigms and raised the problem of incommensurability as a problem of communication between paradigms.

In a development from understanding of concepts as coherent, theory-like and Kuhn's theory of paradigm change as well as Piaget's ideas of knowledge construction, Posner et al. (1982), describe conditions for conceptual change where according to them, there is dissatisfaction with existing ideas, conflict between ideas is perceived, and new ideas appear fruitful and plausible.

The Posner et al. (1982) model has had a profound influence on subsequent theorizing. Scott et al. (1997) describe a range of literature that belongs to this perspective. Others, including Bendixen (2002, p.191-212), Ferrari and Elik (2003, p.21-54), for example, describe the process of intentional conceptual change as 'perturbations' to an individual's concepts that causes the individual to question current understanding. This is said to be followed by deliberate efforts to account for those perturbations. Posner et al. (1982, p.223) conclude that teaching science involves a rational basis for conceptual change. They further argue that conceptual change will be rational if students have the commitments or the standards for judgement required at their disposal.

Gill et al. (2004) argue that conceptual change theory offers a model to understand change in conceptions. They too cite Posner et al. (1982) to suggest, after a consideration of research, that the model is effective in inducing conceptual change in other academic domains including teacher education. Here my argument is with words such as 'induce' considering that change in concepts is more likely to be a volitional and thoughtful act.

My research explores learning as conceptual change as a both psychological and social process. The proposition of a perception of agency as relating to change in task based discourse defines a link between the social context and the perception of agency within it. Learning as individually constructed in social contexts is problematised and the effect of the tension of structure and agency as inherent in social constructivist learning situations is argued. A possible relationship of conceptual change with a learner's perception of agency is hypothesized to contribute to the process of theory building. The review of literature on conceptual change is thus an important part of the study and informs the empirical research.

2.2.4 Understanding Conceptual Change

In keeping with the dual perspectives that co-exist in the domain, my understanding is that concepts and change of concepts are complex phenomenon and processes. Psychological, social and philosophical aspects need to be integrated in analysis. From the viewpoint of psychology, Pintrich (2002) goes so far as to argue that there is no need for research on whether individual factors or social contexts are mechanisms of conceptual change, considering this to be a given. I find that an either or stance on the process of knowledge construction as a function of cognition with its associated role of reflection, *or,* as a social construction is not useful. This study argues concept development as a product of individual cognition within social contexts.

My thesis is built on the premise that a perception of agency within task based discourse provides an individual with space to reflect upon own prior concepts and the new ideas being introduced. The data tentatively supports the hypothesis. As I have noted above and in the conclusion of the research, individuals within the sample changed differently. There is tentative evidence to suggest that more and better change in epistemic beliefs took place in groups where a perception of agency was enhanced. My efforts to urge participants to

agree or disagree, to think and reflect upon their own ideas and experience and those of others could possibly have made a difference.

It may be possible to argue that a perception of *agency was naturally at work for some, to a lesser extent, in both groups,* Control and Experimental. In the Experimental group the perception of agency to reflect upon prior ideas was deliberately enhanced bringing about more and better change. Change did take place in the Control group as well but to a lesser extent.

It is important to remember here that the second aspect under study is the effectiveness of Skemp's proposition that experience of a range of suitable examples can support the development of concepts. This experience was provided to both groups. The data in both groups, Control and Experimental, shows that significant change took place which supports the effectiveness of this strategy. Thus the *active experience* of a range of suitable examples of conflicting accounts on key historical events did serve to bring about change.

Schön, Newman, give importance to contexts, actions and words in developing convergence of meaning. Newman (1999, p.81) disagrees with Schön's claim to have developed an epistemology to account for change and the emergence of novelty. He revisits and analyses Schön's case studies to argue that reflection in action or reciprocal reflection in action are redundant for convergence of meaning. Newman (1999, p.157) claims that convergence of meaning comes about through participation in language games and a reinterpreted reflection may take place afterwards.

I have concerns about Newman's arguments. Convergence of meaning is a problem when there is incommensurability between paradigms as argued by Kuhn. There is commitment to existing paradigms. Conflict is perceived followed by attempts at conflict resolution. Such situations would necessarily provoke thought, reflection and deliberation. An unquestioning assimilation or infusion of incommensurable ideas is difficult to imagine.

Newman discusses a tension in Schön's notion of frames, family resemblances and incommensurability with his positing of sense data 'as a referent giving incorrigibility' Newman (1999, p.35). He describes Schön's proposal for a

dialogue that has three essential features; it takes place in context, it makes use of actions as well as words and reflection in action. Newman disagrees with the need for reflection. He finds Schön's case study 'unremarkable' and his analysis flawed Newman (1999, p.69, 73). He further argues that Schön fails to account for the means by which convergence of meaning is achieved between novice and expert practitioner Newman (1999, p.82, 111).

From a psychological perspective, Piaget, followed by Russell, give importance to agency in mental development. I examine a *perception* of agency within the social contexts of task based discourse to question, agree, disagree and reflect as making a difference in change. My data finds tentative support for this. Certainly, reflection and reciprocal reflection in action did take place in my workshops. Participants did reflect upon and surface their prior epistemic ideas as both responses to questionnaires and excerpts from the focus group discussion illustrate. That they did become aware of their own stance on knowledge in history and did change is evident in the data taken at pre-test and post-test. That they did come to learn, if weakly, the processes of enquiry and knowledge development in history with the use of the examples and the contexts of task based discourse, has empirical support.[13] This may be described by Newman as an initiation into the language games of historians.

Evidence in this research however suggests that participants did reflect in action as they participated in the task based discourse. Without noticing, recognizing conflict with their prior concepts, struggling to see the implications of the evidence, accommodating to the idea that evidence can be questioned in reference to the question posited, change in their epistemic beliefs could not have taken place. The nature of the change, that of epistemic beliefs, is a recognition that knowledge in history is constructed from fragile, discontinuous evidence and the presence of the historian in the account is complex and must necessarily require questioning, requires awareness of conflict and reflection in action. The data reveals individual differences in the degree and quality of change suggesting

[13] See excerpts from the Focus Group discussion on pages 18 and 203.

that prior concepts, experiences, mental capacity did play a role. The evidence also suggests that a perception of agency to reflect and question within task based discourse made a difference in change.

Baltas (2007, p.63) discusses the problem of relativism as inherent in the assumption that there is no extra-paradigmatic neutral ground as Kuhn argues. All meaning residing within paradigms makes paradigm comparison impossible therefore all knowledge can be questioned as relative. Arguing for the possibility for reason to counter the threat to relativism, Baltas (2007) proposes that after the Eureka moment the new paradigm affords a broader, wider perspective. Thus providing space for reason and avoiding the notion that any paradigm is as good as the other. Several participants in my studies demonstrate change from naïve objectivist positions towards an awareness of knowledge as constructed. This raises questions if this is to be called a paradigm change, or a move to a newer, wider perspective.

I digress from here to remind readers that while Kuhn takes a broad perspective on change he too discusses change at the level of the individual with sensitivity and detail. See Section 2.2.10.

2.2.5 Intentional Conceptual Change

Sinatra and Pintrich (2003, p.6) describe intentional conceptual change as ‘goal-directed and conscious initiation and regulation of cognitive, meta-cognitive and motivational processes to bring about a change in knowledge’. They further argue that intentional conceptual change involves some internal agency, volitional control, or self-regulation in the process of learning new concepts. Sinatra and Pintrich (2003).

Bandura’s theory of social cognition and self-regulated learning has taken on many meanings in psychology and education according to Reeve et al. (2008). Reeve et al. (2008), refer to Deci and Ryan, (1985) to explain autonomy as ‘the sense that one's actions emanate from one's self and are one's own’. They cite studies that show that when learning tasks were introduced in ‘autonomy

supportive' as opposed to 'controlling' ways, students showed more positive learning outcomes.

Pintrich et al. (1993), Sinatra and Pintrich (2003, p.ix) famously criticised the Posner et al. (1982) model as 'cold conceptual change' raising awareness of intention and motivation as factors involved in the process of change. There are even examples in the literature of 'hotter', 'warm' or 'warmer', conceptual change models. Motivation and intention are variously considered either external or internal factors.

Vosniadou gives importance to prior learning. In a criticism of radical socio-cultural or situative perspectives, Vosniadou notes that the focus is upon the acquiring of cultural tools, practices, artifacts and argues that the active role of the individual in understanding or constructing new knowledge must be taken into account. She believes that the conceptual change approach could be reframed to include both. Vosniadou (2007a, p.3).

Sinatra and Pintrich (2003, p.380), find that there is agreement that conceptual change does not just take place in individual minds but socio-cultural factors and educational settings also play a role. Inspired by Bereiter and Scardamalia's (1989) idea, of the 'intentional learner', Sinatra and Pintrich introduce their construct of intentional conceptual change to argue that the learner's intentions can determine the likelihood of conceptual change, Sinatra and Pintrich (2003, p.ix). Conceptual change, say Sinatra and Pintrich (2003, p.2), does not just depend on cognitive factors such as recognition of conflict, but on factors of affect, motivation, metacognition, which a learner may control thus governing the possibility of change. They focus on the role of the learner rather than teaching process and aspects such as epistemological stance, self-regulation, meta-cognition, and motivation that would support change. It is important to recall here that Pintrich et al. (1993), earlier suggested four general motivational constructs, goals, values, self-efficacy and control beliefs, as potential mediators of the process of conceptual change.

Vosniadou (2003, p.404) discusses whether intentional learning is indeed needed to promote conceptual change and suggests that more complex change

mechanisms eg model making, would need intention. She considers it is possible that conceptual change can take place without intention but such change may be less stable or inconsistent. Hatano and Inagaki (2003) consider that intentional conceptual change is not necessarily a frequent occurrence. Hatano and Inagaki (2003, p.407). They give importance to instruction including teacher/peer support, activity, internal rewards, the socio-cultural context as well as individual motivation. They argue that a large scale revision is produced only when there is no other choice. This again coheres with Kuhn's descriptions of change in my understanding.

A comparison of intentional and unintentional levels of cognition is made by Sinatra and Pintrich (2003, p.431) as they explain intentional level processing as goal directed, involving some metacognitive or meta conceptual awareness and under the learner's conscious control. This argue Sinatra and Pintrich (2003, p.4) is initiated by the learner. They also refer to some internal agency in the process.

Ferrari and Elik (2003, p.22-49) suggest that any comprehensive account of conceptual change must necessarily integrate both individual agency and culture in its analytic framework. By 'culture' they refer to cultural symbol systems developed over generations. They suggest that persons engage radical conceptual change for concepts where it has personal critical importance. Further that, through the provision of material and educational conditions that nurture intentional conceptual change, we involve students in learning. My research looks beyond material and educational conditions to the social context of groups involved in task based discourse and the agency that individuals perceive within these contexts to develop their own understandings.

2.2.6 Adult Learning in Groups

Adult learning has its own extensive literature. Cognitive control that has been most researched, according to Knowles et al. (2005), is that of field dependence/independence. They urge a distinction between behaviours of self-teaching and the internal cognitive process of feeling and acting with autonomy.

Rogers (2002, p.175-178), analyses the advantages and disadvantages of groups for adult learning. He thinks that 'the pressure to conformity, the suggestibility that the group exerts can promote imitation not the free exercise of experiment'. Members may find it difficult to express a divergent point of view and revert to old ideas when away from the group. This I consider a valuable insight in understanding adult learning in groups and as relating to our observations of learning in teacher education classrooms.

The differences of prior experiences of individuals in learning groups is often taken into account by the teacher in the process of a brain storm to draw out prior knowledge at the beginning of a lesson or within a group activity. How individuals themselves address conflict with prior ideas and the new concepts within the group discourse, and what factors of context can make a difference to change, requires more research. This gap in knowledge is the focus of my study.

2.2.7 The Argument for Agency

Kuhn's (1962) theory of paradigm change is often cited in beginning discussions on conceptual change. Here I repeat for readers what I have stated in the introduction in order to revise and emphasise this. I have argued that the processes of change in paradigms in scientific communities as described by Kuhn could be incommensurable with change through instruction. Instruction has a change agent and planned change is assumed first and then brought into production. Therefore, to import Kuhn's theory to explain and predict conceptual change in classrooms one must recognize and accommodate the difference. The autonomous, cognitive, socio-cultural processes that make existing ideas obsolete and generate consensus and enthusiasm for the new, creating a new paradigm, are not part of planned classroom process. Change through instruction is generated deliberately by others in artificial situations and on possibly unsuspecting actors. Classrooms are places for more 'mediated action' as opposed to the autonomous, naturally evolving action that is taken as examples of paradigm change.

The process of paradigm change, according to Kuhn, involves discovery, the awareness of anomaly, followed by exploration. Crisis, tension, resistance, polarization, mass persuasion, displacement, are words in the text that pertain to the process. Kuhn argues that 'The transfer of allegiance from paradigm to paradigm is a conversion experience that cannot be forced', Kuhn (1962, p.151). After Kuhn and following on from Posner et al. (1982), the structure of dissatisfaction, conflict, plausible and useful alternatives, is often imported into possible classroom experiences it is hoped will mediate change. This is perhaps a reasonable approach but incomplete. Other factors need to be considered as well.

It is in this context of complex learning that Kuhn's analysis of changing paradigms can be revisited to account for change even at small scale and individual levels. The change in paradigms as social-cultural events have been as I highlight, autonomous, often hotly debated, emotionally charged. Issues are real and the change is often radical and especially as described by Kuhn in the domain of science, the change has ultimately been effective over large groups and over time. In classrooms, issues may be of transient, of cursory interest. Participants may not see the need to invest energy in what they may see as an interest of the group rather than a personal concern.

In the Asian - Pakistan context, discourse is more polite, and constrained. There are socially desirable positions and often a lack of understanding is masked by silence. What may also be lacking is agency or perceptions of agency to raise alternative views. The concept discussed in class even most energetically, in the context of active tasks will be constructed in different ways by the same participants outside of group pressure.

Changing concepts is not a straightforward additive process of inserting new knowledge bytes into old frameworks. Change of concepts needs experience of conflict and plausibility of new ideas as framed by Posner et al. (1982). Space is also needed to reflect upon meaning and to accommodate, at will, the ideas on offer.

While the change I observe is in a concept, from a sociological perspective, questions to understand are these. How do teachers-as-learners perceive the

newly introduced ideas? How does the discourse alienate or include participants? Does the dynamic of whole and small group discussion and seeming consensus for new ideas actually mediate against change for some? How does the teacher's role as bearer of new ideas effect change? Elements such as pace, sequence, language, differences in prior experience or prior cognition, could frustrate the adult in a group discussion or task. The learner may want more control and may feel group pressure to conform to the new ideas as coercive.

Change as intentional, assumes that individuals know and clearly understand their existing concepts, the new concepts on offer, and change consciously and intentionally. This also assumes that individuals do not feel constrained by group pressure or discourse to accept or reject the proposed concept or theory. An underlying assumption of social construction is that the social context is, universally, a positive influence on individual conceptual change. That universally, individuals construct own understanding in synchrony with the knowledge constructed by the group. Real life experience, indeed paradigm change as described by Kuhn, does not fit this assumption.

If Kuhn's structure is to be taken to build theory of the mechanism of change, then account needs to be taken of the social conditions that pertain at such a time of change. The tension between agency and structure needs to be seen as part of the equation in individual conceptual change. What this implies is a perception of agency as a variable in group discourse, and agency to be seen as a variable in successful construction.

Group discussions are often set to produce objects reflective of a group view, group knowledge. This is different from situated cognition or actional perspectives which are task or situation related. Legitimate peripheral participation[14] describes the initiation and participation in communities of professional practice. Discourse in classrooms has been endlessly researched. I do not know yet of a term that adequately describes the distributed abstraction that is the product of group discussion, a socially constructed knowledge that may end just there, in the atmosphere, leaving tutors in the false impression that good

[14] Lave and Wenger (1982)

learning has taken place. How does this group viewpoint differ from the knowledge that individuals may have constructed in the same time and space may be worth knowing.

Teachers, as students in workshops, may over a period of time, assimilate the learned ideas within their own cognitive framework. They may embrace the idea wholeheartedly, engage partially or disregard it. This again, however speaks for the space outside of group pressure where a perception of agency brings reflection and conversion to the new idea. If this is indeed the case, then adjustments need to be made to the way the social processes of knowledge construction are conceived towards an awareness of the need of the individual's agency to construct own concepts within the social context.

More so for adults, instruction to change concepts and possibly beliefs, raises issues of consciousness, of empowerment, intention and motivation. Research on conceptual change now focuses on intentionality being a factor.[15] Ethical issues also arise when a possible change cannot be known or understood in advance. Participants, for example, are not competent to judge if they wish to attain greater maturity and 'sophistication' in epistemic cognition. Were they competent to do so they would not need instruction.[16]

Teaching to change concepts is intentional, where teachers are clearly aware of the scope of the goals they plan. How such teaching can be made effective, whilst still humane, moral, not manipulative and not coercive, is a complex issue.

2.2.8 Schön's Epistemology

Conceptual change is, as argued above, a complex phenomenon which is studied from various perspectives. From the perspective of philosophy and teacher education, the work of Schön is a large influence as the common use of terms,

[15] See Ferrari and Elik, Mason, and Vosniadou in Sinatra and Pintrich eds.(2003)

[16] The problem perceived in changing concepts in adult education is illustrated in the account that is presented as a Case Study. (See Appendix A1) An event in which this researcher was a participant is described.

reflection on action and the reflective practitioner, bear witness. I will not go into this in detail here just refer to a small but key aspect that is especially relevant to the objective of conceptual change and the contexts of task-based group discourse that are the focus of this research.

Newman (1999, p.38) is critical of Schön's attempts to answer the crucial epistemological question of convergence of meaning if meanings are internal to paradigms. He raises Schön's[17] argument for context, actions, and words as well as reciprocal reflection-in-action as important conditions for convergence of meaning. Newman (1999, p.81) examines Schön's case studies to contest his empirical observation of convergence of meaning between expert practitioner and novice through a process of reflection in action and reciprocal reflection in action.

Newman (1999, p.149) reinterprets reflection in terms of Wittgenstein as describing certain behaviours within a language game, arguing 'that it is within a language game that it makes sense to doubt, to give reasons, and to make what is ordinarily implicit, explicit'. In a critique of Schön's arguments Newman (1999, p.158) contends that reflective practice may have more than one meaning and that, 'in accounting for convergence of meaning, there is no need to posit as an 'essential feature', reciprocal reflection-in-action'. Schön has however been a large influence on the development of ideas and practices for reflection in action as a means to develop concepts in contexts of professional practice.

From the background of psychology and conceptual change learning, I do consider reflection on our prior concepts as part of the process of assimilation and accommodation of new ideas along with the importance of contexts, actions and words in the learning process. This research is an inquiry into this process.

In this particular research I am interested in better understanding the contexts of task based group discourse and the knowledge that is constructed within these situations. As described earlier, our experience as teacher educators has been that assessment of the products of group learning often does not correspond with the learning of individuals within these groups. Exploring how such

[17] Schön (1987a, pp.100-101) referred to in Newman (1999, p.157)

group learning situations can be made effective in the change and development of concepts of individual members within these groups is useful for my practice.

In my experience of mentoring, I have often found that teachers ascribe the difference between what the group had conveyed in discussion or in group presentations, and pupils' individual concepts in written responses to questions as a problem of conveying ideas in writing. The effectiveness of the task based discourse as a context for individual learning is seldom questioned. These are practical, everyday problems in classroom learning and require research as well as theoretical consideration to resolve.

In response to such problems of practice, I focus my research upon a perception of agency of individuals participating in 'task-based discourse'. This includes contexts, actions and words and reflection as important conditions for convergence of meaning. I consider Newman's arguments against reflection or reciprocal reflection in action being necessary for convergence of meaning in the light of the description of, and empirical evidence of, the change in concepts that I observed in my studies, to draw tentative conclusions as I proceed. Recognizing reflection as part of the broader context of language games raises interesting questions about the nature of such knowledge.

The gap in knowledge about conceptual change in task based discourse is illuminated in my research. The study explores the relationship of an enhanced perception of agency making a difference in task based discourse to persuade more participants to change epistemic concepts. Evidence is seen to tentatively support the relationship. Contexts, actions and words in the use of a range of suitable examples are also seen as effective in developing learning. See Section 2.2.10 for more on reflection.

2.2.9 Agency and Mental Development from a Perspective of Cognitive Psychology

Russell (1996) discusses the psychological and philosophical nature of the arguments raised in understanding the processes of cognitive development with

reference to Piaget (1928). Importantly, Russell (1996) explores the strength of a theory of agency driving cognitive development from infancy to maturation. Piaget's theorizing, although based upon research on cognitive development of children, is however, a large influence in the understanding of the development of cognition in general. It is important to note here that the nature of development under consideration is the development of ' the process of establishing a division between two kinds of reality: an objective reality grasped as independent of ourselves and a subjective reality constituted by our volitions and representational states' Russell (1995, p.127). Russell explains agency as the capacity for first-person experience, Russell (1996, p.2).

A claim by Piaget that agency is the *engine* of mental development is recognized by Russell to be untenable. (italics mine) He finds that there is too much evidence of innate capacity to accept such a claim. One example Russell (1995, p.129) employs for innate capacity is of research by O'Keefe and Nadel[18] that finds animals possess innate, within the hippocampus, capacity for spatial cognition. While accepting the criticism against Piaget, Russell argues for a watered down version for the role of agency in mental development. It is better to cite his own words here.

'My principal claim here will be that it is only by experiencing agency that a subject can experience the world as resistant to her will, something which is necessary for any distinction to be drawn between subjective and objective. I argue here that an essential feature of agency is a capacity for *willfully* determining the sequence of one's perceptual inputs', Russell (1995, p.127). (Italics mine) He also argues that the contextualist option cum gradual dawning is open, even preferable for us, Russell (1996, p.205).

There is a fine difference between the proposition that a perception of agency relates to change in concepts in task based group discourse, which is the hypothesis I propose for my research, and Russell's arguments for the central role of agency in mental development. My hypothesis relates a perception of agency to change in concepts within the social context of task based group discourse. This

[18] Russell refers to O'Keefe and Nadel (1978)

throws light on the function of individual agency within group discourse involved in a learning task. Context, action, and words are present but the social context needs attention.

This argument does not challenge the widely accepted notion of knowledge as socially constructed or of learning as a development of shared meanings in social interaction. Theories of learning in social interaction such as Legitimate Peripheral Participation in Communities of Practice, Lave and Wenger (1999) or the Community of Inquiry framework, Garrison et al. (2000) give importance to context and participation in learning in activity. Within this approach issues are raised of power relations and conditions for legitimacy as defining possibilities of learning, Lave and Wenger (1999). Garrison et al. (2000) focus on collaborative learning and consider a ‘cognitive presence’ as 'the extent to which learners are able to construct and confirm meaning through sustained reflection and discourse' Garrison et al. (2001).

These theories do not directly concern themselves with the process of learning new concepts or psychological aspects such as assimilation or accommodation of new ideas at an individual level. Knowledge is in participation.

Prior concepts are individual properties and change within these must take place at the level of the individual even when in contexts of participation in active tasks, discussion and reasoning in social groups. Products of group learning activity in terms of responses may not always reflect the concepts of all individuals within groups as teachers often observe. Alternative ideas held by individual learners are seen to surface when teachers inquire.

A perception of agency as a condition in task-based group discourse that can relate to individual conceptual change is the focus of this study. The hypothesis that agency relates to change in task based group discourse does not imply either that learning best takes place in isolation. What is argued is only that there are constraints within task based group discourse for individuals to construct their own concepts, making connections with personal prior knowledge and experiences. If change in concepts needs to overcome the tension between assimilation and accommodation at the level of the individual, the task based

discourse needs to be supportive of agency to question, to reflect, to agree or disagree, to argue, to speak one's word. The constraints in group discourse are of a practical nature as well as at a level of process of cognition. The space to reflect and make meaning at an individual level, where after all, concepts have to develop, is restricted in social situations and may contribute to resistance to change. A perception of agency is explored in my research experiments as making a difference to learning within task based group discourse.

The understanding that knowledge is not simply constructed, it is co-constructed has its basis in the theories of Vygotsky and extended in work such as that of Lave and Wenger (2002)[19] on knowledge as situated in communities of practice. This is not challenged in my study. I do consider that there is a tension in the function of the construction of an individual reality in the presence of multiple realities making the role of a perception agency all the more crucial. By enhancing a perception of agency in the social situations of group learning as a pedagogical device, I suggest change in concepts may be better supported.

The meaning I give to agency acquires a fine nuance. Agency as perceived by individuals in social contexts of group discourse adds the dimension of perceived freedom to alter one's perceptual inputs to what may be own prior knowledge without the necessity to uncritically assimilate the conceptions of others. The tension between assimilation and accommodation arises in contexts of new experience, and agency being a necessary condition, is necessary due to the particular prior conceptions of the agent. As an agent, at the level of the individual level, agency is a necessary condition as I can assimilate or accommodate only to the conceptions I already possess in my own mind and not to those of others.

2.2.10 Reflection in the Process of Conceptual Change

My arguments for a perception of agency in task based discourse emerge from an analysis of Kuhn's illustrations of paradigm change in communities of

[19] Lave and Wenger in Harrison et al. (eds) 2002.

scientists. I do very briefly raise Schön, Newman here as questions may be asked regarding the relevance of Schön to my own arguments or the learning that took place.

Newman (1999) discusses Schön and Wittgenstein's work and finds support for Schön's argument for the importance of contexts, actions as well as words, in developing a convergence of meaning. Newman (1999, p.113) however, points out that while Wittgenstein considers actions and contexts are important for the convergence of meaning, reflection in action has already been explored by him (Wittgenstein) and found 'incoherent'[20]. Newman is critical of Schön's claim of 'reflection on reflection-in-action, reflection-in-action and knowing-in-action[21]' considering these notions to be 'redundant'[22] for convergence of meaning.

The pedagogy employed in the development of concepts in my research workshops uses aspects of context, action and words to good effect. Here I draw attention to the fact that the situation in my workshops was not the one studied by Schön of apprentice and expert in an exploratory dialogue within a practice context. Teachers as learners participated on tasks in small groups in the workshops. They took part in actions of examining historical sources, identifying conflicting accounts and used heuristics to argue a claim. There was task-based discussion amongst participants themselves, and with me the expert, as they considered their prior concepts and as the evidence suggests, they did in small individual ways, review and change their epistemic beliefs about knowledge in history. From paradigmatic beliefs in knowledge in history as objectively known, to not possible to know, or to a more criterialist stance, change did take place to various degrees at the time.

If meanings are internal to paradigms, how can convergence of meaning take place is the essential epistemic question. I do draw attention to the observation that change, or convergence of meaning took place within the learning experience of the workshops in my study. There were no other experiences as

[20] Newman (1999, p.113)
[21] Newman (1999, p.113)
[22] Newman (1999, p.113)

possible causes of development of epistemic beliefs operating within that time and space. Apart from myself, none of the participants in the small groups were experts. Evidence suggests that participants' epistemic beliefs changed, which is qualitatively different from simple knowledge accretion. The experience itself was novel for participants as they explored historical evidence and attempted to construct accounts. They tried to argue and justify these on the basis of the strength of the evidence. The discourse required an exploratory use of epistemic ideas, the language of source and evidence, certainty and justification with which they were not familiar as the excerpts at the beginning of this Chapter and others, illustrate. At no time did I or the students consciously and overtly raise or discuss philosophy or epistemic beliefs as phenomenon underpinning the tasks they carried out but we discussed truth claims, evidence and heuristics whereby accounts could be examined. To discuss philosophical theories would have been difficult to say the least as the participants did not know the words or were familiar with, the same language game.

Did reflection support the change in beliefs or was it the broad context of the language games exploring the concepts of knowledge in history that they experienced that made the difference? This is a question that could be raised here. As I describe in this study, there was modeling, discussion, questioning and the tentative use of words and methods of analysis of evidence that historians use. The questions, as I point out, were of an epistemic nature and ones the participants were not familiar with. Certainly individuals must have reflected at the time to be aware of the anomalies between their own prior concepts and the new ideas. The new ideas did appear plausible as the small excerpt at the beginning of the Chapter illustrates.

It is important to note here that change in participants' beliefs was not even or consistent within groups. It took place in various ways and to varying degrees. The evidence suggests that participants' beliefs changed from those they came with, in generally the direction of more sophistication according to the CBKH scale. The differences in their prior understanding as well as their personal capacities

that I discuss with reference to Vosniadou and Hatano[23] possibly affected the degree and quality of change amongst them. The responses to the questions show a range of meanings in the use of words, upon the interpretation of which, I draw conclusions as I assess the change that took place from pre-test to post-test.[24]

Kuhn's description of the processes of change from one paradigm to another are broad in scope, discussing change at the level of the society of scientists within relevant time frames. It is useful to note that Kuhn focuses upon, explores and describes, often in detail, difficulties in change at the level of individuals such as Priestly, Lavosier, within the larger complex of change in the community as instances of change. Reisch (2012, p.324) refers to these as case studies. One such small reference[25] to Lord Kelvin's reaction to the discovery of X-rays to be an elaborate hoax illustrates the possibly intense and painful, personal reaction to the challenge. When Kuhn (1970, p.83-84) further describes the response to crises, he refers to Wolfgang Pauli, who, when faced with what was to be a new quantum theory, said he wished he had been a movie comedian and had never heard of physics. In my sample of teachers from different school systems across the country I can see that, broadly, change did take place within the community as they participated in the language games of knowledge in history. Yet it is not homogenous or equal at the level of individuals. Interestingly, as the change is in epistemic beliefs and within the direction and language of the change in the group it does demonstrate coherence.

In reference to my central claim, I find meaningful where Newman (1999, p.45) refers to Schön….'advocates of conflicting frames (to) enter into one another's worlds, attempting under a willing suspension of disbelief, to learn the things, the coherence, the other has created there.'[26] A close reading of the above highlights the aspect of 'willingness'. For the purpose of argument, I wonder if

[23] See page 34 (in this Dissertation)

[24] See 3.36-3.38 for an understanding of the beliefs and quality of change

[25] Kuhn (1962,1970, p.59), 'The Structure of Scientific Revolutions' in International Encyclopedia of Unified Science

[26] Newman (1999, p.189) has provided a comprehensive list of the work of Schön which is useful to find references.

‘willing suspension of belief’, is a condition that would support an enhanced perception of agency?

From Newman’s point of view it could be questioned if participation and development within a language game brought convergence of meaning if, as he argues, reflection is redundant and “it is within a language game that it makes sense to doubt, to give reasons, and to make what is ordinarily implicit, explicit’ Newman (1999). In my study I provide a role and time for reason and reflection within discourse and activity My research focus is upon a relationship of a perception of agency within task based discourse to reflect upon own prior concepts in order to develop these concepts. For this I find tentative support.

2.2.11 Reflection on My Own Learning and Epistemic Belief Change

Prior to the beginning of this research, I was a teacher, educator and an editor of a small school teacher research journal. I organised much of the research myself. Looking back at my own work, I can see that I did not at that time understand in depth what I read and even taught, about the nature of knowledge as constructed. My own beliefs were naive and my awareness of the meaning of interpretation and construction of meaning was instrumental rather than philosophical. Newman (1999) quotes Schön, ‘once practitioners notice that they actively construct the reality of their practice and become aware of the variety of frames available to them...’ This *noticing* is a giant step and it is in essence the one necessary in epistemic belief change. I have argued a beginning ‘awareness’ as in a sense a loss of innocence in the development of epistemic concepts. This ‘noticing’ or ‘awareness’ can lead to a change in paradigm and, as research into epistemic beliefs documents as well as my own developing understanding tells me, it does sometimes take place. In Section 2.6.7, I discuss awareness of knowledge as constructed as the defining difference between an objectivist and subjectivist stance on the CBKH scale. I argue that finding participants had not as yet developed expert skill in the use of words and heuristics but were aware that

knowledge claims were constructed needed to be unraveled to place them on a scale of epistemic development.

2.2.12 Skemp's Principles for the Formation of Concepts

Conceptual change in mathematical concepts is argued by Skemp (1971) in detail in his text The Psychology of Learning Mathematics. Skemp considers that concepts are not learnt by definition but by encountering, experiencing, a suitable collection of examples. This I have amended to an 'experience of a range of suitable examples' and used it as a pedagogical method in my studies. The examples are described in the Chapter on Research Methodology. The use of a single pedagogical method in all 10 groups helped to reduce the threat of alternative inference of cause. As change was significant from pre-test to post-test in both Control and Experimental groups, the data supports the effectiveness of Skemp's principle in bringing about conceptual change even in this case in epistemic beliefs.

Skemp (1971, p.32) suggests that it is the communicator who must know the principles and indeed, I as teacher, communicator, endorse his view. From the perspective of psychology, Skemp (1971, p.29) includes, abstracting, classifying, naming, communicating as part of the process of developing concepts and argues that, 'The actual construction of a conceptual system is something which each individual has to do for himself'.

2.2.13 Reasons for a Focus on Epistemic Beliefs and Beliefs About Knowledge in History

Upon embarking on a study of change in concepts, I required a concept to change within the teacher education sphere I was involved in and I found this in the domain of history where a change was in process. The history syllabus that students in Pakistan study for CIE, Cambridge International Examinations, is undergoing development apparently to bring it in line with international standards.

History teaching and learning has moved away from an exclusive focus on first order concepts to include development of epistemic beliefs about knowledge in history. The teacher education project that is the context of my research could thus centre on developing teachers' concepts about knowledge in history in keeping with the requirements of syllabus change.

History education in the developed world includes aspects of historiography. Ashby et al. (2005) state that if history education is to include an understanding of history as a discipline, then central to that understanding is a concept of evidence. Ashby et al. (2005, p.24), Maggioni et al. (2009b) also see important connections between historical thinking and the development of historical beliefs.

VanSledright (2011, p.1) addresses the 'knowledge teachers need to possess in order to significantly deepen their students' historical understandings…' locating this in epistemological underpinnings as well as other second order concepts. This is a very real concern as changes in syllabi can be met in superficial terms as well as active resistance when teachers recognize their own inability to understand and deal with the new ideas. Thus, the rationale for my research and the workshops that were the context is reinforced. Anecdotal evidence gathered during the pilot and in the experimental workshops found that most teachers understood that the sources provided in the test papers were merely meant to be hints for students to help form the answers to questions of history.

VanSledright and Afflerbach (2005, p.1) argue that knowing something about the nature of a source helps situate it within an array of different types of historical evidence that can be used to build an interpretation of the events under consideration. They argue that assessing the status of sources is the sine qua non of historical understanding because access to the past is 'largely, perhaps, solely, mediated by evidentiary accounts and artifacts'. Evaluating source and evidence, criticising and developing accounts are part of the process skills of history and have been a part of progressive curricula over the last 20 years. CIE has now begun to introduce the change in the Syllabus of Pakistan Studies. How this develops and its impact could be the subject of research studies.

2.2.14 The Importance of Research into the Development of Concepts About Knowledge in History

History teaching and learning more than any other subject area, is a matter of concern from a range of perspectives. Apart from mainly political debates about the content of history, arguments about the purpose that history education serves are raised. VanSledright (2002, p.4) expresses concern that importance placed on literal comprehension of text during reading activities develops a belief that the 'meaning is in the text, that the text contains what really is'. He describes how children use text books where the historian author is not apparent thus producing a belief in the 'reality effect', that all the words in the text map directly onto what is real. Teachers and tests that require verbatim recall also reinforce such ideas he argues. If this is the finding in America, history text books in Pakistan are no better. The textbooks produced and used in Pakistan are often geared to reduce, summarize and focus content to set questions and most national examination of history require tightly circumscribed responses rather than argument as a critical review of questions repeated over the years in textbooks and papers can show. Needless to say this has a backwash effect on concepts of knowledge as well as learning and teaching practices. The planned change in the CIE Syllabus towards a focus on history skills is much needed to encourage development of concepts of knowledge. My research into teacher's epistemic beliefs and change within these is an effort in this direction.

Learner's epistemic beliefs about knowledge in history are a source of concern. Bain[27] is often cited in the literature for describing history as fundamentally an 'epistemic act'. Referring to Barthes (1968, 1986), VanSledright (2002) discusses 'the referential illusion' explaining this as the belief amongst readers that interpretations of historians literally mirror a past reality. Barton (2008) cite Ashby and Lee (1991) writing that 'children, adolescents and adults often approach sources as decontextualised, disembodied, authorless forms of neutral

[27] Bain (2000: 3-4). See for example, VanSledright et al. (2006: 560)

information that appear to fall out of the sky ready-made'. I agree with Bain and consider the content of text books that teach history as facts for rote learning and repetition as harmful in developing critical thinking. Ashby et al. (2005, p.80) find that students have critical misconceptions that have an impact on the way they read and imagine the past. Making history accounts problematic by raising epistemic questions is at the very purpose of study in the discipline, argues Bain (2005, p.179-211).

Barton (2008) reviews research on students' ideas about history and the topics they study at school to find that most students do not recognise history accounts to be developed from evidence. Barton (2008, p.240) finds that even if students learn about sources they may not be critical. Wineburg (1991) observed that adults when asked to write their story starting from available evidence, tended to reaffirm the objectivity of the text by seeking to be absent from their own narrative. Maggioni et al. (2004, p.176) argue that an attitude recognizing positionality will seldom be cultivated if the role of the historian is not recognized since 'unchecked present standards are often used as a default to interpret the historical evidence'.

While difficulties in change are described in the literature there are some encouraging findings. A study conducted by VanSledright (2002) drew attention to the role of instruction in challenging students' epistemological stances and to the difficulties inherent in this endeavour Maggioni et al. (2004, p.177). VanSledright (2002) did find, what he calls, conditional success in helping fifth graders engage in efforts to think about history although Maggioni et al. (2004, p.176) in a review of studies on historical thinking found that adults did changed little when presented with a new understanding of what history might be.

My research into conceptual change targets teachers' epistemic beliefs in the domain of history. I believe I made a difference in developing the epistemic beliefs of the teachers who participated in my study. The study has also helped me personally, in developing my own understanding and beliefs about knowledge. Can we, as teacher educators, bring about this change needed for teachers and through them their students? My research may help in answering this question.

Barton and Levstik (2004, p.191-260) argue for inquiry teaching considering that inquiry makes the process of knowledge construction transparent although they recognize problems in the method.

2.2.15 Concerns and Purposes of Teaching History

There is a concern to be recognized in teaching history as constructed. As Maggioni et al. (2004, p.186, 191) discuss, novice students exposed to conflicting evidence without being provided the tools to deal with the tension may jump to 'resigned naive relativism' or 'cynical skepticism', concluding that history is just a matter of opinion. I, too have concerns that when development from an objectivist stance to awareness of the nature of knowledge in history as constructed takes place, there may be possibilities that learners may be left at that stage without progress towards a more criterialist stance. The danger of relativism, that all accounts are true accounts or that there is no way to judge between them, requires that educators leading learners on this journey see it to its more informed and educated end. My efforts did change teacher's beliefs from pre-test to post-test. It would be interesting to see what epistemic beliefs teachers continue to hold or develop after a passage of time.

2.3 Review of Literature for Research Methodology

2.3.1 Arguments on Paradigms and Mixed Methods

The design of my research based on quasi-experiments with the integrated use of both qualitative and quantitative measures as well as methods of analysis, was formulated after much study and thought to frame reasonable answers to study questions. These questions arose naturally in the peculiar context of interest in the relationship of conceptual change with perceptions of agency amongst adult learners. Contextual opportunities and constraints also influenced the type of design and means of data collection.

Creswell and Plano Clark (2011, p.38) suggest that mixed methods researchers should not only be aware of their philosophical assumptions but also clearly articulate these. Articulating ones assumptions is easier said than done as these are slippery things and often context related in my experience. However, as this research is situated in a delicate balance between critical realism that is the philosophy of quasi experimentation as claimed by Cook and Campbell (1979), and the process of interpretation of epistemic beliefs from tentative statements to open ended questions about knowledge in history, there is a requirement to explore theoretical perspectives and a possible stance in order to build some kind of order to the exercise.

In this Chapter, I explore philosophical perspectives that are said to underpin mixed methods research in general and my own developing ideas in particular. Pragmatism is the philosophical perspective that is often used in arguments for mixed methods. Many of the decisions I have made in my research are pragmatic and practical. While I can support aspects of pragmatism that allow space for researchers to address complex research issues and questions, I find employing personal values as a basis of choice between explanations of cause interesting but requiring some thought. I question the possibility of using personal values to choose between competing explanations as needing definition and instead of 'what works' I also consider 'what is known to work' as support for design choices as there is a connotation in the latter of what is known to work in the academic community. 'What is known to work', as an evolving concept, would be more responsive to knowledge growth, the dynamics of the concept, and the diverse and tentative nature of knowledge in history. The research itself could demonstrate how mixed methods work in practice**.** Upon reflection upon the conduct of this research, I reframe my stance. This is discussed further.

The use of mixed methods in research has its critics as well as supporters. Arguments on philosophical perspectives underlying mixed methods research have been daunting to say the least. It became important to study the various points of view in this domain, develop insight, and make an informed choice. The research design is experimental although set in the field, therefore arguments for

mixed methods need thought. What follows is a review of some important perspectives in the debate and how these informed a choice of mixed methods for this quasi experimental research.

Teddlie and Tashakkori (2009, p.77) firmly state that researchers working within the mixed methods tradition may find certain aspects of 'strong' postmodernism difficult to reconcile with the very act of performing research. They cite Gorard and Taylor (2004) that by denying the possibility that there is any means of judging knowledge claims to be more or less true, postmodernism makes research a completely pointless activity. To quote Teddlie and Tashakkori (2009, p.77), 'Among strong postmodernists the gap between the collapse of objectivity and science is taken as beyond serious debate, and inquiry is judged by humanistic…..standards of aesthetics, poetics, morals, and interpretive insight, rather than by objective standards of truth'. This strong criticism can become an argument for a critical realist or pragmatist stance.

Guba and Lincoln (1994, p.109), who describe themselves as transcendental realists, argue that differences in paradigm assumptions cannot be dismissed, implicitly or explicitly. These positions they say, have important consequences for the practical conduct of inquiry as well as for the interpretation of findings and policy choices Guba and Lincoln (1994, p.107-112). Guba and Lincoln define paradigms as basic beliefs, worldviews that have to be accepted on faith. They argue that paradigms are human constructions and advocates of any particular construction must rely on persuasiveness and utility rather than proof in argument. Bazeley (2002, p.3) considers that one can't research or prove paradigms, and paradigmatic debates can never be resolved.

Howe (1988, p.15) considers that theories of truth as correspondence or truth as coherence are a forced choice as do Teddlie and Tashakkori (2009). Howe is critical of the two positions, positivism, championed as truly 'scientific' and interpretivism, embracing multiple realities with researchers free to speak their own languages, investigate their own questions, and come up with their own standards of truth. Howe (1988) argues that the two paradigms, positivist and interpretivist, do not exhaust the possibilities and reminds readers of pragmatism as a possible

alternative. Howe and Eisenhart (1990, p.2) argue that the problem of standards for inquiry is best framed in the 'logics in use', which I tentatively apply in practical terms. When examining 166 responses to 6 open ended questions, it was logical to develop a rubric that made a common standard of assessment possible and transforming the descriptors into numerical scores for analysis.

Smith and Heshusius (1986, p.6), in an often cited article in the qualitative–quantitative debate, criticise attempts by Guba and Lincoln to develop criteria to judge quality in naturalistic research. Smith and Heshusius (1986) argue that Guba and Lincoln's view together assumptions underpinning naturalistic inquiry alongside criteria and procedures that characterize rationalist inquiry. They criticise LeCompte and Goetz (1982) as well, who, they say, describe procedures for the practice of naturalistic inquiry in ways that are comparable to rationalistic inquiry, making their approach as systematic and rigorous as the other. Miles and Huberman (1994), argue Smith and Heshusius, ignore paradigmatic differences in order to pay attention to developing a body of clearly defined methods for drawing valid meaning from qualitative data.

Smith and Heshusius are critical of Miles and Huberman who according to them do seem to accept that concepts such as valid, real, dependable, trustworthy can be defined in the same way for both sides. The principal concern now is, complain Smith and Heshusius, to find methods for qualitative inquiry that will provide the same objectivity and certitude presumably available to quantitative inquiry. They argue that as reality is mind-dependent, a description can only be matched to other descriptions and not to an 'unconceptualised reality' Smith and Heshusius (1986, p.7).

For qualitative inquiry, Smith and Heshusius (1986, p.8) argue, a valid interpretation is among interpretations one with which 'one agrees'. They elaborate 'the ultimate basis for such agreement is that the interpreters share, or come to share after an open dialogue and justification, similar values and interests.' Qualitative inquiry, insist Smith and Heshusius, does not require that certain things must be done or that validity is a matter of appropriately implemented techniques. All that can be done, they say, is to match descriptions to other descriptions,

choosing to honour some as valid, given one's interests and purposes. The last adds a radical pragmatic explication in Smith & Heshusius's argument. The founding arguments are epistemological and ontological but do not resolve on this basis an epistemologically grounded answer to the dilemma posed. The choice between alternative interpretations, choosing to honour one, given one's interests, is therefore, pragmatic and personal.

From my perspective as an educational practitioner and end user of knowledge, one whose responsibility is to change practice, this is problematic. In education, the responsibility of children's learning and indeed their life chances, makes educational decisions a matter not so much as one of personal choice as *responsible* choice. It is questionable if the stance of Smith and Heshusius is coherent with the stance on knowledge in normal educational practice. Schools are places where assessment and evaluation of learning and teaching are everyday activities that have consequences for pupils, teachers and schools. Personal values as a basis for choice between knowledge claims are not the norm. The alternative to personal choice however, of claim of knowledge corresponding to reality can be, and has been, successfully challenged on many fronts. Problems of learning and teaching need knowledge that works to resolve problems and build content for the development of practice. It is reasonable to assume that truth would work in practice while that which suits personal or group interest may or may not.

Teddlie and Tashakkori (2009, p.90-91) explain their pragmatic stance distancing themselves from post-positivists on axiological considerations. They say pragmatists decide what they want to study, and include units of analysis and variables based on what is important within their personal value system. Thus far, pragmatism is a practical and useful philosophy.

Discussing paradigm issues in mixed methods research Teddlie and Tashakkori (2009, p.87-89) argue that pragmatism and transformative perspectives reject forced choices between positivism/post-positivism and constructivism with regards to methods, logic, and epistemology, and embrace features associated with both. Teddlie and Tashakkori (2009, p.87-89) believe that pragmatists and transformative scholars can choose both inductive and deductive

logic in the research cycle to address questions. This theoretical support helped in planning my research including both confirmatory and exploratory questions as well as both pre-set categories and grounded development of categories in analyzing data.

Teddlie and Tashakkori (2009, p.90) discuss the relationship between knower and known on a continuum rather than on two poles. They suggest the researcher and participants may require a highly interactive relationship at times and distance at others. They address ontological considerations, pragmatically suggesting that they agree with an external reality but deny that truth regarding reality can actually be determined. Pragmatists, they say, are unsure if any explanation of reality is better than another. They cite Howe (1988) that for pragmatists, truth is a normative concept. Howe explains that for pragmatists 'truth is what works', is best seen not as a theory or definition, but as the pragmatists *attempt to say something interesting* about the nature of truth and to suggest, in particular, that knowledge claims cannot be completely abstracted from contingent beliefs, interests, and projects (italics theirs). Transformative scholars suggest Teddlie and Tashakkori (2009, p.93) believe there are diverse viewpoints and choose between alternative explanations that best promote social justice for oppressed groups. The last is a choice based on personal morals which again sets aside arguments for truth.

Teddlie and Tashakkori (2009) say they believe that research conducted in the tradition of pragmatism is carried out within the value system of researchers and is focused upon answering questions in the researchers interest. There is a crucial difference between the two statements. The first refers to choice between explanations on the basis of personal values, the second to design elements which is probably the norm for most research. If values are to be taken to refer to philosophical, disciplinary or cultural viewpoints or positions, my question is: would these then be the values of the individual or the community of scholars and reflect a democratic aspect? Individual or small group values in the global community are hard to defend as what may be political and ethical values for one community may not be so for another. Pragmatism, as the philosophy avowed by historians,

especially those of American History, can thus be challenged by marginalized groups in ways that are similar to the History of the Sub-Continent written by those from other dominant groups. Kloppenburg (1996, p.101) express a concern that present day pragmatists differ from early pragmatists who believed their philosophical ideas had particular ethical and political consequences. They find that contemporary pragmatists consider it merely a method of analysis.

This is where I part ways with core pragmatism. To dwell at length on philosophy, while fascinating, is outside the scope of this research. Suffice to say that to choose pragmatism alone as an expedient alternative conceptual framework, for me, would be unwise. I, as teacher educator, Pakistani, female, Muslim, student researcher, inferring cause on the basis of personal values, would be opening myself to argument quite distinct from that of research quality. Symonds and Gorard (2008, p.1) argue that labels can be restrictive in enforcing categorical differences. They suggest as a practical alternative, to focus on the quality of research techniques, data, and how that data is used instead of trying to construct overarching categories and researcher identity Symonds and Gorard (2008, p.17). This stance is more pragmatic, less controversial, and more in line with my personal values. One last rationale is the practical choice in a given situation such as the many I have taken in the quasi experimental design.

Howe and Eisenhart (1990, p.3) argue that abandoning positivism does not lead to abandoning standards. Howe, as a pragmatist's argument, is that as there is no way to know truth, truth is what works. Howe (1988, p.15) argues that knowledge claims cannot be completely abstracted from underpinning contingent beliefs, interests and projects. Pragmatists, he argues, giving the examples of Kuhn (1977), Quinne (1970), supplant coherence and correspondence with criteria such as accuracy, scope, simplicity, consistency, and comprehensiveness Howe (1988, p.15). I will need to explore what this means in practical terms.

Shadish et al. (2002, p.35) explain pragmatism in a different light. They suggest that pragmatism says a claim is true if it is useful to believe the claim which is somewhat different and more practical from a claim that suits personal values. They cite Latour (1987) that what comes to be accepted as true in science is what

scientists can convince others to use, what they say is implied in Mishler's (1990) assertion that qualitative methods are validated by a 'a functional criterion - whether findings are relied upon for further work'. Shadish et al. (2002) reason that as all theories of truth, correspondence, coherence, pragmatism are compromised, practicing scientists should not have to choose between them in justifying a viable approach to the validity of inferences about causation and its generalization. This is a position I consider useful in the conduct of this research in teacher education using a quasi-experimental and mixed methods design. I have arguments however, with core pragmatism which I discuss at various points in this dissertation.

Pragmatism's choice of 'what works' may not be enough justification for some. But pragmatism as *what is known to work* could be based on experience and expertise. Smith and Heshusius (1986, p.9) make even this problematic by arguing that the problem with what works, no matter how it is expressed, tells us nothing about the process of inquiry and the interpretation of results.

Shadish and Cook (2009) who are proponents of quasi-experimental research discuss theories of truth and suggest that as all roads to truth are compromised, their theory of validity employs all three, correspondence, coherence, and pragmatism. They argue that scientists should not have to choose between them. Perhaps so, however, to be unaware of such concerns especially using mixed methods would be a recipe for trouble.

Where I select a set of open ended questions as a data gathering instrument, I do so because this affords respondents space to construct a broader, personal, reflective response. I select Likert Scales where I require more tightly circumscribed responses taking a stance. When I transform qualitative descriptors to quantitative scores it is a practical choice to facilitate analysis of a large number of responses. I am pragmatic as a practical choice in context but I resist labeling the decision as a personal value.

As far as drawing conclusions using both qualitative and quantitative data and methods of analysis, I am content to be careful in applying method and principles prescribed for each methodology in its place such as those of internal

validity, being transparent and providing rich description, drawing tentative conclusions within the limitations that are found, and being open to argument. I do not expect that a stance identified on the Likert Scale would be the self-same stance expressed in the open ended responses or one a more true isomorphic, reflection of the reality of the respondents' beliefs than the other. I do however expect to see some coherence or what is referred to as an 'overall fit'.

In the absence of consensus on paradigms for research, mixed methods properly applied, and carefully conducted are at present the best choice. Method can gain ascendancy of careful rule following, taking context into consideration, and being open to criticism. *What is known is of value then, only if it is known how it is known, and to what parameters it can be held to apply.* In the real world, knowledge is needed and used for real purposes to address urgent real world problems. Just as products carry labels of contents and ingredients, etcetera, research too should have openness, detail, including if any, a statement of the personal value that has given rise to the choice. Apart from this, a critical, reflective awareness of background assumptions at every stage could make knowledge claims more credible. A way forward for future research could be to see if assumptions constructed to personal values do indeed work to solve real world problems.

Being a mixed methodologist, I have found, is being beset by self-doubt and indeed, doubts about theorists that seek to light the way. It took courage to realize that pragmatism is the way forward especially in complex areas. Like all theories however, it is better to take pragmatism itself, critically, as a work in progress, that practical research activities can help to direct and define.

Here I am reminded of a phrase in the Quran which I shall take the liberty to quote, 'Reality! What is the Reality? Ah, what will convey unto thee what the reality is!'[28] Indeed it is difficult to take a stance to answer this question especially if one is a social scientist. The vagaries of fortune of theories even in the hard sciences make this challenge applicable to all knowledge claims.

[28] (Chapter 69: Verses 1,2,3) Translation by Muhammad Marmaduke Pickthall

This quasi-experimental and mixed methods design straddles the spheres of experimental science with interpretive approaches and finds this comfortable and practical in achieving the research purposes. Theoretical arguments are necessary in academic work. This research demonstrates in its own way, how ideas and actions can be made coherent to achieve a real life educational goal to find out, to make what we teachers do, better.

2.3.2 Effect Sizes

The results are tantalizing and interesting but not conclusive in the sense of statistical significance in most tests of the relationship. This not being a randomized control trial but a practitioner research using mixed methods, results are interpreted with care. Numbers being small in the complete population (83), and within single studies, and selection being non-random, limits the number of tests that can be conducted as well as the strength of the interpretation. There are no easy references of effect on adult learning therefore the references for children's learning are used. It is important to remember that epistemic beliefs and change within these, as well as a perception of agency are psychological constructs unlike self-report grades, homework or feedback which can be expected to have a larger impact on children's learning as Hattie's Rankings lists show[29]. Effectiveness of the intervention is seen with a test of percentage change on all items of the BLTHQ. Cohen d effect size is 0.511.

There is no effect size quoted in Hattie Rankings (September 2014, May 2015) on a perception of agency as this may be the first time that this relationship has been suggested, researched, or reported, to my knowledge. The sizes quoted for related psychological constructs such as 0.48 for motivation, 0.24 affective attributes, are taken as a rough reference. Thus, effect sizes ranging between 0.51 and 0.60 in tests, as seen in data in this research, can be interpreted from good to moderate. Some tests show significant difference.

[29] Self-Report Grades 1.44; Feedback: 0.73. Hattie Rankings May 2015

Hatties Rankings of effect size fall in the genre of ‘what works’ research in school-based learning therefore these are not easily generalised to teacher education. Hattie (2009, p.7-8) quotes Cohen (1988) that an effect size of 1.0 is ‘large, blatantly obvious’. It is useful however to know that an effect size of over 0.40 is considered 'hinge point’.[30]

2.4 Arguments for the Design of the Quasi Experiment and Validity of Findings in Quasi-experimentation with Reference to this Research

As stated, my research design is based on a Quasi Experiment in the context of in-service teacher education in schools where I have been a practitioner. I have relied mainly upon the texts of Cook and Campbell (1979) and Shadish, Cook and Campbell, influential theorists in the field, to design the Quasi Experiment. Campbell and Stanley (1963) introduced the terms Quasi Experimentation, and Internal and External Validity. Rosenthal and Rosnow (2008), and many of their arguments focus on these aspects. This Section of the Chapter discusses theoretical aspects relating to internal validity while its practical implementation is discussed in Research Methodology.

In this Section, I review and discuss the theoretical assumptions that underpin my own project. I have found that ground realities often reflected much of what was described by the authors. In a sense, generalising theory to the particular contexts of research in schools in Pakistan did become possible. I see a pragmatic perspective, a leaning towards 'what works' underpinning the reasoning of Cook and Campbell (1979) and Shadish et al. (2002) as they describe quasi experimentation. The authors also say they are ambivalent about their role in promoting quasi experimentation and there are increased efforts on their part to emphasise better control in selection to make causal inference more reliable. In more recent work, Shadish and Cook (2009) say that Randomised Control Trials are neither necessary nor sufficient for well-warranted causal inference.

[30] Grant Wiggins as read on 7th May 2015: grantwiggins.wordpress.com/2012/01/07/what-works-in-education-hatties-list-of-the-greatest-effects-and-why-it-matters/

Cook et al. (2010) find that the theory and practice of field experimentation is increasingly being used for answering causal questions. The term 'quasi' is described in the Concise Oxford Dictionary as meaning seemingly, apparently, but not really, and that is how I have found quasi experiments to be in reference to true experiments.

Cook and Campbell (1979, p.35), Shadish et al. (2002), outline that in quasi-experiments, the cause is manipulable and occurs before the effect is measured and that the paradigmatic assertion in causal relationships is that manipulation of a cause will result in manipulation of an effect. The experiments in my research were designed carefully to keep the cause manipulable and occurring before I measured the effect.

Cook and Campbell (1979, p.6-8) describe quasi experiments as experiments that have treatments, outcome measures and experimental units but do not use random assignment. Thus Experimental and Control Groups may be *non-equivalent* where field conditions constrain possibilities of randomisation. These non-equivalent Groups, warn Cook and Campbell, may differ from each other in many ways other than treatment effects which could be alternative explanations for the observed effect. Cook and Campbell (1979, p.6-8) however, argue that random assignment makes most of the alternatives less likely as causes of observed effects. Importantly, Cook and Campbell (1979) and Shadish et al. (2002, p.xvi) emphasise a focus on better design rather than dependence on statistics to test causal propositions. Their descriptions of various types of experimental designs as well as design elements that help to counter threats to generalised inference were valuable in planning. Threats to validity are referred to as heuristic devices in examining the research design. No list is exhaustive and I have found that I need to be alert to threats that may be peculiar to my research context.

2.5 Truth Claims in Quasi Experimental Research

Cook and Campbell (1979), as self-described ‘critical realists’, tread the ground with care as they develop their logic of experimentation. They are careful

to state that ‘outside variables will always impinge on a dependent variable making results sensitive to forces outside the theoretical system’ and, that they expect the relationships to be fallible and probable at best. Cook and Campbell (1979, p.15-34) use the term validity to refer to ‘the approximate truth of an inference’ and lay out procedures that they suggest can support assumptions of internal construct, and external validity in making claims regarding causal relationships between variables. Cook and Campbell (1979), Shadish et al. (2002, p.34) recommend modesty and use of the terms ‘tentative’ and ‘approximate’ in making truth claims as ‘one can never know what is true’. Taking from Popper, Cook and Campbell say that at best one can only know what has not yet been ruled out as false (Cook and Campbell (1979).

Shadish et al. (2002, p.34) say their theory of validity makes use of each of these approaches to truth; correspondence, coherence, pragmatism, but recognise that each is compromised.

2.5.1 The Assignment Process

In my programme of research, I chose what Cook and Campbell (1979) recommend as generally interpretable, Control and Experimental Groups with pre-tests and post-tests design. The programme had five separate studies in three different school systems over three cities of Pakistan. Each study had a Control and an Experimental group. In all studies, selection of participants or assignment of participants to Control or Experimental groups was not mine as the researcher, as described above. Groups were formed naturally, based on geographical location. My own decision was arbitrary and across the board; the second of two groups in a study, would be the Experimental Group as described above.

Can this process be seen as having the probability value of a chance assignment? *As the assignment process was unrelated to the outcome, on average, could this be seen to function like a randomised experiment?* That is a matter of judgment. What appears to me to be more meaningful is that the groups

were typical of those found in teacher educational classrooms with their multiple similarities and differences.

Steiner et al. (2009) find that within study comparisons have shown that quasi-experiments regularly fail to reproduce experimental results unless the assignment mechanism into treatment is completely known or extensively and reliably measured. Meaning thereby that with knowledge of the assignment mechanism, the study could be replicated. I can only guess the assignment mechanism and certainly not measure it in reliable terms, but I assume that it would on average be in a similar range in all groups with individual differences. I consider, knowing schools well, that on average in a group, there would be a similar range of reasons why participants chose to attend or chose the Control or Experimental Group, especially as they did not have knowledge of the difference in the treatment in the groups.

Those designs that meet the standard for inference include non-equivalent control group designs with plausible theories of selection into treatment versus control states and extensive and reliable measurements of this selection process Steiner et al. (2009). The last clause would necessarily preclude most quasi-experimental designs including this one. As described in the last Section, it was found to be difficult to obtain reliable data regarding the selection or assignment process therefore ‘extensive and reliable’ measurements of the process were not attempted. Many key questions such as prior qualifications in history appeared threatening to the participants and responses were either withheld or not reliably answered.

Insisting on responses would have been a threat in itself and conflict with the perception of agency I wished to create in the experimental groups. These are the field conditions that make true experiments impossible and even restrict possibilities of inference in quasi experiments. By making this shortcoming plain and avoiding interpretation that is more than tentative, I hope to make this research useful and credible. Other practitioners trying out a perception of agency in their classrooms would however, benefit from knowing that conditions were similar to those in their own contexts.

Steiner et al. (2009) write from the context of research in psychology, especially university based research, where randomised trials may be feasible. Students in psychology departments would be more likely to cooperate in providing information regarding selection, and in random selection. Administrators and managers in busy, children's schools would be unlikely to cooperate in random selection and assigning staff to groups. These are the realities of the field. There is relevant information regarding participants that would remain hard to obtain or unreliable if volunteered in many cases.

2.5.2 Validity

Validity, argue Shadish et al. (2002, p.34), is a property of inference not a property of design or method. Campbell (1957) explained internal validity as the question ‘did in fact the experimental stimulus make some difference in this specific instance?’ The data appears to show that in 4 out of 5 studies the experimental stimulus could have made a difference. How valid or reliable is this inference depends upon what is construed as valid or reliable in such quasi experimental conditions.

Shadish et al. (2002, p.38) list four types of validity; statistical, internal, construct, and external. Cook and Campbell (1979) describe a range of threats to internal validity and encourage the use of design to control these threats where possible. Cook and Campbell (1979) point out however, that this is a deductive process and fallible.

Shadish et al. (2002, p.63) see a close relationship between internal and statistical conclusion validity. Both, they say, are concerned with study operations and with the relationship between treatment and outcome. Shadish et al. (2002) discuss construct validity and raise the question, ‘how can we generalise from a sample of instances and the data patterns associated with them to the target constructs they represent?’ This is a hard question to answer but their suggestions are reasonable.

A perception of agency, the treatment, is a construct that may be interpreted in various ways. Therefore, I chose to limit the definition to specific actions and statements which I could follow. The outcome, change in epistemic beliefs, especially as observed in responses to the open ended questions, again required clear explication to make assessment consistent. The effort to develop the rubric, CBKH, and repeated assessment, as well as Inter Rater and Internal Evaluations was useful in keeping a careful eye on the process. Using the same three different instruments to measure change and carefully following classroom procedures and content kept all studies as similar as possible and supported the level of inference that could be made.

Shadish et al. (2002, p.69) extend the usage of construct validity from outcomes and treatments to include persons and settings. They argue that construct validity involves making inferences from *any* of the sampling particulars in a study to the higher order constructs they represent (emphasis theirs). Understandably, they warn against construct mislabeling. This is a concern of mine as the measures I use, imply a sophisticated or naive set of epistemic beliefs, which can be open to challenge. Critics, respondents, can challenge the validity of a conclusion. I have used these existing measures as a matter of convenience as any measure of such concepts would be an approximation. Sampling treatments, say Shadish et al. (2002), is rarely done in field research for good practical reasons and I did not as well. I piloted four instruments and did not use one, The Four Quadrant Scale, Schraw and Olafson (2008, p.25), as it appeared to have social desirability effects in my sample.

2.5.3 Fuzzy Plausibility

Shadish et al. (2002, p.484) argue that quasi experiments rely on researcher judgments about assumptions, especially on the fuzzy but indispensable concept of plausibility. While both texts describe the possibility of quasi-experiments, a concern for improving inference through careful design is emphasised. Shadish et al. (2002, p.484), suggest that in the best of quasi

experiments, internal validity is not much worse than with the randomised experiment and argue that fuzzy plausibility is central to ruling out threats in quasi experimentation. After following arguments to overcome threats to validity in the research, I found I could counter several threats but not all. Key among these of course is random selection and assignment; nor was I able to measure and account for threats as key information was simply not available. These I have raised and described giving reasons. This begs the question that with these reservations, why are quasi-experiments considered at all? The answer, as I have found, lies in both the constraints of field conditions where random selection is near impossible as well as the poor relevance of true experimental conditions to the real life world of schools.

2.5.4 Types of Designs

Cook and Campbell (1979) discuss two classes of designs; those they consider generally non-interpretable, that generally do not permit causal inference, and non-equivalent, control group designs that, they say, are generally interpretable. Among the latter, the untreated control group with pre-test and post-test is most often used and often interpretable. Cook and Campbell (1979, p.103). This is the design I have used in all 5 studies. Groups were not matched in particulars. Broadly, all belonged to the category of O' Level teachers of history in high school. They are therefore classed as nonequivalent.

Shadish et al. (2002) recommend Regression Discontinuity designs where the decision to assign participants to Control or Experimental Groups is based on a measure prior to treatment. Assignment to treatment is on the basis of a cutoff score on an assignment variable not by coin toss or lottery. Interestingly, Shadish et al. (2002, p.216) say that the assignment variable can even be one that is totally unrelated to outcome and have no particular substantive meaning. They cite the example of Cain (1975), where the order of application to a program was used as a basis for assignment. The first 20 applicants were assigned to treatment, the remainder were controls. Such a rule did not remain constant in my study as

participants self-selected to groups according to location. The two groups formed for each study were assigned a location at a distance and participants actually came to the group nearer to home or school.

Another example discussed is that of Deluse (1999) where couples were assigned to treatment based on the day they filed for divorce. This, argue Shadish et al. (2002), would function like a randomised experiment in which the assignment process is unrelated to the outcome on average. I find this an interesting argument as it can be extended to apply in a wider range where assignment is to criteria that do not impact the outcome. In my 5 Studies I kept an arbitrary rule to decide which group would be the Control and which would be Experimental. The first of two groups in each study was the Control Group and I kept this practice throughout the 5 studies. Thus this is an assignment variable that is totally unrelated to outcome. For me this has to suffice.

2.5.5 Selection Bias

Steiner et al. (2009) refer to research (in carefully circumscribed contexts and with regression discontinuity designs) that suggests bias reduction has been achieved with full knowledge or extensive measurement of the selection process and, when intact, local but nonequivalent comparison groups are selected that heavily overlap with the treatment group. In other contexts, they argue, the degree of bias reduction has been disappointing.

The groups in this research are not intact but formed in all cases at random in the sense that no particular selection process was employed. I as researcher, was presented the groups as found by schools. Participants could have either been self-selected or assigned by school managers. I am known as a teacher educator of long standing, and there is some awareness of the value of training for CIE programmes amongst high school teachers and schools, therefore the workshops could have attracted interest and those available could have elected to come. These are conjectures however. In many ways the groups were representative of such a group formed in normal teacher training workshops. Some information was

volunteered on the forms but it was not adequate in developing sound knowledge of participants' qualifications and experience. Probing questions were seen as unwelcome, therefore I gave up asking. My focus on a perception of agency made me nervous of creating undue tension or awareness that could possibly create differences in treatment.

Cook and Campbell (1979) and Shadish et al. (2002) warn against what they call mono operation bias and mono method bias. Using several measures can address the former as I have done by using both quantitative and qualitative measures. I have also used a measure of general epistemic beliefs as well as two of particular history related beliefs about knowledge to extend the description as well as for possibilities of triangulation. I have, however, used the same set of measures throughout the five studies in order to facilitate comparison. I have conducted the experiment in different locations what Shadish et al. (2002) call multi-site studies. Different persons from a range of schools and cities were used in experiments. While broadly falling into similar categories, finer differences such as age, years of experience, etcetera are the norm.

2.5.6 Generalized Causal Inference

While discussing generalised causal inference, Shadish et al. (2002, p.385), mention that researchers are often interested in the level of treatment implementation that would be anticipated as modal if the intervention were routinely adopted. Indeed, in this case, building a perception of agency would have become ineffective if it had become noticed as a peculiar action, a research treatment, or a comical aberration in the trainer's behaviour. The need to make the actions less obtrusive did however clash with the problem that participants may simply miss seeing the message if it was too lightly presented. This is another kind of threat. In any follow up research, I would still recommend that sincerity in the perception of agency would be more effective than one where it was seen merely as a gimmick.

Shadish et al. (2002, p.96) cite Cronbach (1982) that timely, representative but less rigorous studies can lead to reasonable causal inference while Campbell and Boruch (1975) are said to maintain that causal inference outside of experiments is problematic because many threats to internal validity remain unexamined and have to be ruled out by fiat rather than through design and measurement.

2.5.7 Possibilities of Randomization and Standardization

Shadish et al. (2002, p.23) recognise that random sampling is always desirable but only rarely and contingently feasible. They suggest that purposive sampling is sometimes useful and more commonly used.

Shadish et al. (2002, p.50) cite Boruch & Gomez, 1977; Cook, Habib, Settersten, Shagle, & Degirmencioglu, (1999), Lipsey, (1990), to argue that conclusions about co-variation will be affected if treatment is implemented inconsistently from site to site or from person to person within sites. This threat they say is pervasive in field experiments, however they add that a lack of standardization is intrinsic to some real world interventions. While mechanical aspects were easy to deliver consistently, I cannot claim that all my actions were entirely consistent or were consistently perceived by all. The perception of agency was provided as planned, however in any classroom it would be difficult to claim that all heard, observed, registered, and constructed the same message. This not being a medicinal pill taken at set time times and quantities, there are bound to be variations in the perception amongst participants.

I do not have adequate evidence to make a claim but generally I consider that a difference in experience of the perception-as-operationalized was surely provided between control and experimental groups. I tried to make an audio record of all sessions but the quality of the record varies, being incomplete and inaudible at times, makes it problematic for use. I have used my records and the audio record in order to describe the experiments and to illustrate the process. For a replication of the experiment, in depth qualitative data would be needed as well as more

thorough observation. These are difficult to manage in self-funded research as well as being impractical in terms of all participants.

The treatment was not complex and did not involve actions on the part of others which made it easier to monitor. It did, however, involve care in delivery so that all introductions and statements were made and the follow up approach was persisted with throughout the three day course. The treatment was novel. In such a case Shadish et al. (2002) recommend that levels of treatment should be included; one as high as possible and one equivalent to a no treatment group. This I will need to recommend to others who wish to conduct this research.

It must be kept in mind that field research has dynamics that can change with the imposition of greater demands. More measures can cause boredom and attrition. My most pressing concern at all times was that of participant attrition due to the content being perceived as less than relevant. Such fears dominated the list of influences on the design and conduct of the research. Therefore, while the ideal design of an experiment may be desirable for interpretation, what one can actually do in school is quite another matter.

2.5.8 Cause in Quasi Experimentation

Cook and Campbell (1979, p.10-15) want to base causal inference on procedures that reduce uncertainty about causal connections. Steiner et al. (2009) understand that full explanation of any causal relationship is necessarily context dependent and, further, that many factors are required for a given cause-effect relationship to occur.

Shadish et al. (2002, p.9) argue that while experiments are useful for *causal description* they do less well in *causal explanation*, which is clarifying the mechanisms through which, and the conditions under which, that causal relationship holds (Italics added here).

Shadish et al. (2002, p.10) argue that causal explanation is an important route to the *generalization* of causal descriptions because it tells us which features of the causal relationship are essential to transfer to other situations. I recognize

the modest possibilities of generalization but have added descriptions supported with contextual data from field notes, records, photographs, and random observations by participants.

Shadish et al. (2002, p.27) understand that the impossibility of theory neutral observation…..implies the results of any single test (and so any single experiment) are inevitably ambiguous. They further suggest that such critique is true of single studies and less true of programs of research, Shadish et al. (2002, p.28). I took great pains to repeat the experiment in 5 studies in order to add to numbers in a meta-analysis as well as to study the relationship in broadly similar yet different in particularities of persons/settings/times, to see how well the relationship between agency and conceptual change holds. This is an aspect which will be reviewed in analysis.

2.5.9 Alternative Explanations in Causal Inference

Shadish et al. (2002, p.14) explain that in quasi-experiments, the researcher has to enumerate alternative explanations one by one, decide which are plausible, and then use logic, design, and measurement to assess whether each one is operating in a way that might explain any observed effect, Shadish et al. (2002). This is a process I have followed although it may be hard for readers who do not know quasi-experimentations theories of internal validity to understand why each threat must be considered and described. I have realised that writing up a quasi-experiment in accordance with the theory of internal validity as argued by Cook and Campbell (1979) requires a very context related, explanatory and descriptive mode which I can assume is different from the writing of true experiments. These transparent details and acknowledgements however, must help teachers and educators recognise familiar territory and the possibility of research within it.

On the other hand, Shadish et al. (2002, p.16) argue that it is neither feasible nor desirable to rule out all possible alternative interpretations of a causal relationship and suggest that plausible alternatives be the major focus. Shadish et al. (2002, p.15-16) argue that falsification depends on two assumptions that can

never be fully tested; the claim needs to be fully specified and require measures that are perfectly valid reflections of the theory being tested. These are indeed difficult claims to maintain in most social science.

In case threats to validity cannot be ruled out by design controls, either because the logic of design control does not apply or because practical constraints prevent available controls from being used, Shadish et al. (2002, p.40) suggest three critical questions. 1) How would the threat apply in this case? 2) Is there evidence that the threat is plausible rather than just possible? 3) Does the threat operate in the same direction? I have made attempts to identify possible threats and counter these in the design. These could be identified and described in my study context but could not be quantitatively assessed due to a lack of information. This is described in detail in the Chapter on Research Methodology.

2.5.10 Thoughts on the Review of Literature on Quasi-Experimentation

The processes of science as well as theories of science, having been developed over time and systematic reasoning, the importance of existing knowledge in the field cannot be overlooked or undervalued. The literature review was an invaluable exercise in both design and its analysis. Carrying out research itself tests such knowledge and advances it with reason and reflection if only in miniscule ways at times. The elegance of quasi-experimental theory was experienced in the design and conduct of this research as many described situations and relationships came to life.

The theory is easy to criticize and cynical reviewers may see many weaknesses in the perseverance that it takes to continue with the experiment in the face of threats to valid inference that are really apparent and obvious. What makes the quasi-experiment worth doing, in my understanding, is the possibility it affords of researching the practice of schools within real life school contexts. Where internal validity becomes a cause of concern, careful design can to a large extent help to sidestep and may avoid threats. One learns how manipulating design features can actually help improve the quality of inference. There is a

degree of challenge as well as pure excitement in conducting such studies which may not be the purpose but it certainly makes science well worth doing.

Cook and Campbell's work has been criticized for its emphasis on internal validity, and their text for its long and often tiresome detail. This in fact is its very strength. The often arduous reading pays off with better design and more arguable inference. Both texts have been my companions throughout my research and I owe the authors much gratitude. The texts are liberating in a sense, because they legitimize research in field contexts. Quasi experimentation in the field makes real life situations valid experimental contexts and procedures within it assume meaningful value.

I have set the quasi experiment in a mixed methods research design and have found that it works well. Both quantitative and qualitative data and its analysis have a logical place within an experiment as I am able to demonstrate.

Shadish et al. (2002, p.18) find that the strength of experimentation is its ability to illuminate causal inference. There are others such as Chomsky who are critical. I pondered Chomsky's (1959) criticism cited in Cohen et al. (2007) that a singular problem of behaviourism is our inability to infer causes from behavior to identify the stimulus that has brought about the response. I interpreted the criticism as a need for a more expanded approach, providing rich description, using both evidence and argument. On the whole, I consider that teachers and teacher educators can and should try strategies and approaches in quasi-experimental fashion in order to better inform their practice.

2.6 Theoretical Background of Measures and the CBKH Rubric

This Chapter discusses key aspects in the development of the theoretical model and instruments used in this study. Change in concepts and a perception of agency are discussed in Sections 2.1 to 2.2.10. Epistemic beliefs are argued next, followed by a review of literature on research methodology including quasi experimentation. What follows in the next Section is a focus on the measures used and the development of the CBKH rubric.

Apart from the three measures, the BM questions, the BLTHQ and EBI Likert Scales, the CBKH rubric is discussed in this section. The Categories of Beliefs about Knowledge in History are developed and used in this study based upon two previous models, Kuhn and Weinstock (2002) and Maggioni et al. (2009a). This set of categories is used to examine and order epistemic beliefs about knowledge in history on a scale of 1 to 7 from naive to sophisticated beliefs.

The background of the development of the CBKH is first described briefly. This is followed by arguments specific to the development of the CBKH on the basis of theoretical perspectives studied as well as some reflection in action during the assessment of responses to the BM questions. Concerns expressed by the sample of teachers about the problems of history texts and how their pupils debate these in classrooms, is another basis for argument.

A category scale such as the CBKH needs to be recognized as a tool in progress of development, used only to make ends meet until a better one comes along. The CBKH is not a prescription of how epistemic beliefs need to be, nor even a complete description of how they are. The complexity of the phenomenon defies complete description. The literature has been read critically, the various options available studied, and a rubric worked out to fit as best as possible with the nuances of the beliefs I could read in the data. Thus theorizing in the literature, my own concepts and skills, discussions with Maggioni, and the concerns raised by the teachers both inform and limit the development of the categories.

2.6.1 Theoretical Models of Epistemic Beliefs that Inform the Development of the BLTHQ, Beliefs about Learning and Teaching History Questionnaire and the CBKH, Categories of Beliefs about Knowledge in History

Maggioni et al. (2009a), Maggioni et al. (2004) argue the development of the Beliefs about Learning and Teaching History Questionnaire, the BLTHQ, and the model of epistemic cognition, the Copier, Borrower, and Criterialist categories, based on theoretical, methodological and pedagogical reasons. They explain that

they tested if their hypothesis was reflective of various philosophical perspectives but also compatible with data obtained of different groups of individuals.

Maggioni et al. (2004) describe epistemological beliefs in the context of cognitive development in History and argue that their research findings parallel changes observed by King and Kitchener (1994), Kuhn and Weinstock (2002); Schraw et al. (2002). The work of Kuhn and Weinstock, King and Kitchener, Lee and Shemilt, is the main reference for the work on the original categories designed and argued by Maggioni et al. (2004); Maggioni et al. (2009a). These frameworks are discussed briefly below as they are pertinent to arguments for the CBKH as well.

Kuhn and Weinstock (2002, p.125) conceptualize the development of epistemological understanding as the coordination of the subjective and objective dimensions of knowing. Their 'Levels of Epistemological Understanding' order development from what the authors term, realist, absolutist, multiplist, and evaluativist levels. This is described as on a developmental trajectory from children as young as three years old to a final level which is 'most likely never to be achieved'. Interestingly, Kuhn and Weinstock introduce a 'constructivist theory of mind' by middle to late childhood. It is important to note here that stage related development of beliefs is argued in the literature. Chandler et al. (2002, p.146) find competing claims at 'wildly different ages'. Kuhn and Weinstock consider that the transition from a multiplist to an evaluativist level as difficult describing this as the move from a belief in all claims being equally valid reflections of their owners subjectivist perspectives, to the evaluativist's reintegration of objectivity into knowing on the basis of criteria to judge between claims to have more merit than others. Kuhn and Weinstock (2002, p.126). The reintegration of objectivity at an evaluativist level in the domain of history requires a careful and reflective approach as I have found and argue in these pages.

King and Kitchener (2002) describe epistemic cognition as the cognitive process enabling individuals to consider the criteria, limits, and certainty of knowing. The empirically developed Reflective Judgment model of King and Kitchener, suggests that individuals engage in three levels of cognitive processing.

At the first level called 'cognition' individuals engage in processes like computing, memorizing, reading and perceiving. At the second level, they engage in meta-cognitive processing to monitor their progress and on the third level they operate in what is called epistemic cognition where individuals engage in considering the limits of knowing, the certainty of knowing and the criteria for knowing. Hofer (2002, p.6) finds the model encompasses the skills of critical thinking and personal epistemology. I can see an interesting order in the development described by King and Kitchener (2002) reminiscent of early levels of understanding in Bloom's taxonomy at the first level and Schön's reflective thinking at the second level. Epistemic cognition appears to be at a sequentially higher level of thinking in this model which is reasonable.

King and Kitchener (2002, p.38) describe Reflective Judgment as based on Dewey's observation that reflective thinking is called for when people recognize that some problems cannot be solved with certainty. Their development of reflective thinking is described as seven distinct but developmentally related sets of assumptions about the process of knowing which they further reduce to three. These are termed the pre-reflective, quasi-reflective and reflective stages. King and Kitchener (2002, p.39) argue that at the pre-reflective stage, knowledge is gained through the word of an authority figure or through first-hand information rather than through evaluation of evidence. Such people are said to see knowledge as certain and absolutely correct. King and Kitchener describe quasi-reflective thinking as a stage when people recognize that knowledge claims contain elements of uncertainty, which they attribute to missing information or to methods of obtaining the evidence. People at the stage of reflective thinking are described as accepting that knowledge claims are not made with certainty but are 'not immobilized by it' but make judgments of which they are 'reasonably certain', that are 'most reasonable' on the grounds of their evaluation of available data.

It is worthwhile here to discuss an older model of epistemic beliefs in the history domain. Seixas (1996, p.770) explains Shemilt (1987) four stage hierarchy

for analyzing adolescents' ideas about historical evidence. The initial level accepts accounts with no questions about the authenticity or reliability of the source and no question about 'using' evidence other than as information. At the next level, questions about reliability or of sources are asked though still *failing* to *use* a source in the revision of an account. The third level is an understanding of evidence as a basis for inference about the past and finally, understanding the historicity of all historical accounts. The last is an argument for more sophistication, necessitating the revision and discounting of all accounts. This hierarchy is clearly ordered according to ideas or concepts about evidence at the same time relating to what one *does* with evidence to make inferences about the past. The importance of the last point cannot be ignored. I have argued in the development of the CBKH Categories that to be a criterialist, the awareness of the knower must not be lost even when the historical method is used with expertise.

I have highlighted the word 'failing' to compare this with Maggioni et al. (2004) use of the word 'deemed' in the same context. There is a difference in reasoning and volition between the two words which is interesting and either may be representative of a respondent's level of epistemic belief or merely a level of performance in a given case. Where Maggioni et al. (2004, p.188) write for a Borrower, 'In this case, the constructed nature of history is acknowledged but the historical method is not *deemed* to be an effective tool to deal with problems of conflicting evidence', a considered judgment is indicated. There is a difference between a person simply not knowing how to evaluate evidence or not believing that evidence can within reason resolve problems of choosing between explanations in history. The complexity of epistemic development makes for no easy answers.

2.6.2 The Model of Copier, Borrower, and Criterialist Categories of Epistemic Beliefs of Knowledge in History

Authors Maggioni, VanSledright and Alexander, describe the development of the model, the Copier, Borrower, Criterialist, with reference to the Reflective Judgment Model (RJM) of King and Kitchener (2002) and Kuhn and Weinstock (2002) conceptualization of mature epistemological understanding as the 'coordination of the subjective and objective dimensions of knowing', Maggioni et al. (2004, p.173). They say the project draws on three main frameworks used in the study of epistemic cognition; King and Kitchener (2002), Kuhn and Weinstock (2002) and Schommer (1990), Maggioni et al. (2009b, p.191)

Maggioni et al. (2009b, p.194-195) discuss the 6 levels of progression of pupils' ideas of evidence identified by Lee and Shemilt (2003), and find interesting similarities with domain general paths of epistemic cognition. They describe the Copier stance as parallel with the pre-reflective thinking in the RJM and the realist or absolutist level in Lee and Shemilt (2003) progression. Describing the Naive Realist position they argue that it is characterized by a belief in a perfect correspondence between the past and history. Maggioni et al. (2004, p.188) cite Alexander (2003) to suggest that Naive Realist beliefs tend to be typical of novices who still do not possess well developed knowledge of the principles and methods that characterize a specific domain. VanSledright refers to reader's beliefs that history accounts fall from the sky readymade. For those who are not familiar with epistemic issues in the history domain, this sometimes seems strange as names of authors are recognized on book covers. Someone did ask me this question. However, there is consistent evidence even in my data that adult respondents can be seen as equating the past with history, for example, in response to the question, 'what is history?' several responses were, 'the past' or 'reflections of the past'. Maggioni et al. (2009a, p.209) argue, in the case of the Copier, evidence fogs the awareness of the interpretive process at work. Here, Shemilt (1987)'s definition of the highest level retaining understanding of ' the historicity of all historical accounts' is again to be emphasised.

Maggioni et al. (2009a) liken the Borrower stance to the 3rd and 4th stages of Lee and Shemilt's progression, the multiplist level of the LEU or quasi-reflective thinking of the RJM. They name it the Borrower in order to 'highlight that individuals tend to borrow their stories from accounts...on the basis of instinctive preferences or casual selections and do not yet fully know the disciplinary tools used by historians'.

The third level of Maggioni et al. (2009a) is the Criterialist which they associate to reflective thinking in the LEU and the evaluativist model of Kuhn. The authors are specific in the association of greater sophistication with the understanding of heuristics. The Criterialist, as they describe, can ask questions that sources of evidence were not specifically designed to answer. They refer to other heuristics such as considering the perspective of the source and corroborating evidence across different sources to overcome the problem of bias in testimony considering that this makes history possible again. To this point they add reasoning about the past in its historical context to make the approach powerful. There are implications in this for history education of teachers as well as students. In explaining the CBKH categories I argue for the continued awareness of the knower and known at the level of the Criterialist beyond the development of knowledge and skill in the use of heuristics. A focus on heuristics alone may obscure the interpretive process at work and the role of the historian in the text.

2.6.3 Are Epistemic Beliefs Domain General or Domain Specific Beliefs?

Duell and Schommer-Aikins (2001, p.420) argue that measures draw heavily upon the theory upon which they rest and therefore, measure different aspects of epistemic beliefs. There are arguments for domain generality as well as specificity. Paulsen and Wells (1998) report differences in beliefs in students from hard-soft-pure-applied sciences. Pintrich (2002) states that substantial differences persisting in epistemic beliefs across domains were unlikely. The EBI measures domain general beliefs whilst the BLTHQ is used to measure epistemic beliefs in history set in aspects of pedagogical practice. Apart from pedagogical practice, the

BLTHQ is also underpinned by theoretical considerations of what is knowledge in history as opposed to more domain general beliefs. The BM is a qualitative measure of epistemic beliefs about knowledge in history set in more abstract, theoretical terms. The theory underpinning the BLTHQ and the BM with reference to beliefs in history is compatible. Using the BLTHQ and the BM as well as the EBI in both pre-test and post-test to assess change provided measurements of change in beliefs in different aspects of epistemic beliefs and both qualitative and quantitative data. The BLTHQ additionally incorporates aspects of pedagogical practice of teaching history.

Hofer (2002, p.11) refers to much work done that holds to epistemic beliefs being domain general and wonders if these can be seen as related in disciplines. An overview of instruments generally suggests a focus on epistemic aspects of what can or cannot be known, how people can come to know including source of knowledge such as authority, and how certain can the knowing be. A measure of general epistemic beliefs, the EBI and Maggioni et al's BLTHQ are used in this research to measure change in concepts post intervention. A key difference between the two is the integration of the concepts and methods of historiography as defining progression in beliefs in the BLTHQ. Concepts of the nature of knowledge in history and the justification of this knowledge with the pragmatic use of evidence and argument are expected to grow with greater sophistication in epistemic beliefs in history.

Some authors such as Schraw et al. (2002, p.261) include an examination of theory of learning aspects such as speed of learning and the ability to acquire knowledge. Belenky et al. (1986) Ways of Knowing refers to silence, received knowledge, subjective knowledge as well as procedural and constructivist knowing. Duell and Schommer-Aikins (2001). This differs from my use of the term in the CBKH as I use 'Ways of Knowing' more in the procedural or strategic sense. This I took as a practical step to unravel the epistemic belief in the responses from the weak knowledge and skill in use of heuristics. This became necessary in the context of the numbers of responses and making a reliable assessment of beliefs within my sample.

Hofer and Bendixen (2012, p.233) consider that while early models have suggested domain generality of epistemic beliefs there are studies that suggest that there is evidence that beliefs may be domain specific as well.

2.6.4 Does Culture Have an Effect on Epistemic Beliefs?

Effects of cultural backgrounds on epistemic beliefs are a focus of research as are perspectives in history. Epstein (1998) is cited by Barton (2008, p.247)[31] to have found contradictory perspectives amongst students of European American and African American backgrounds. Alexander and Dochy (1995) inspected implicit theories of adults from the United States and Europe about knowing and believing and found repeated evidence that cultural background and educational experience shade the way that adults conceive knowing and believing. The authors argue that we cannot assume that those who speak of knowing and believing are communicating from a shared semantic base. Interestingly, Alexander and Dochy (1995) observed that with less educational training or with less expertise respondents were more likely to relay their views on knowing and believing with less tentativeness and greater conviction. A study of epistemic beliefs amongst secular Jews, religious Jews and Bedouins in Israel reported by Tabak and Weinstock (2008, p.177) found Bedouins had higher percentages of absolutist beliefs in all domains particularly values.

I have used the BLTHQ in a very different cultural context albeit similar in terms of a school and teaching and learning history background for reasons that I have explained in the Chapter on Research Methodology.

My study does not make comparisons of stance as seen in the BLTHQ of teachers in Pakistan and those described by Maggioni, VanSledright and Alexander, or explore these on the basis of cultural differences. That would be outside the scope of my research besides having epistemic issues regarding different paradigms and convergence of meaning in my view. I do describe specific features of curriculum experience that could impact beliefs but these are contextual

[31] In Levstik and Tyson (Eds) (2008) Handbook of Research in Social Studies Education.

aspects and not analyzed in the research. I conducted a Factor Analysis to test the performance of the BLTHQ in my sample and if there were differences in pre-test and post-test in the factors that emerged. The Factor Analysis conducted of data from the BLTHQ in my research has interesting differences as well as similarities with data described by Maggioni, et al. I provide the Factor Analysis but it is outside the scope of my research to compare differences and argue this on the basis of culture as stated above. Word limits also restrain such discussion.

2.6.5 Epistemic Beliefs in the Domain of Knowledge in History and How These Differ from Domain General Epistemic Beliefs

Epistemological beliefs about knowledge in history are considered to have an essential domain related relationship of epistemic beliefs about knowledge in history with the principles and practices of history as these have been employed by historians over time. A brief review of literature is necessary to understand the rationale for this.

Ashby (2004) refers to what (Rogers, 1978, p.7) calls the symbiotic relationship between 'knowing that' and 'knowing how' in history. Ashby (2004) explains that history asks different questions about the past, and it is the individual and specific nature of these questions that determine what counts as evidence in validating a claim to knowledge. This key principle is raised repeatedly in the literature.[32] Seixas (1996, p.766) argues that all problems, including the question which account should we believe, addressed by historical thinkers both novices and experts, are rooted ultimately in pervasive historical traces or in historical accounts. VanSledright (2002, p.6) emphasises that historical thinking requires a complex regimen of investigative techniques. VanSledright (2009) uses the terms 'procedural or strategic knowledge' when referring to methods. VanSledright and Afflerbach (2005, p.16) explain the requirements to move students to greater cognitive sophistication. They emphasize the importance of criteria investigators

[32] VanSledright (2002), Maggioni, VanSledright, Alexander (2009), Maggioni, Alexander and VanSledright (2004), Lee & Shemilt (2003), Ashby (2004).

use to assess the reliability of sources. They elaborate these as a move from binary distinctions, towards an awareness of the author's historicized positions. Reading subtext, corroborating testimony, analyzing coherence in accounts, and resolving conflict with interpretations that make the most sense after a careful assessment of evidence are some things students would be expected to learn.

Background knowledge of the historian is seen as a key element. Seixas (1996, p.770) relates the historical thinker's background knowledge of the situation and about the perspectives of the author of the account or record, leads to the revision of the account and to new interpretations. This is what teachers often describe as taking place in classrooms.

It is interesting to read the difference in the use of the heuristic as underpinned by the concept of its purpose by novice and expert. Finding the students and experts not just performing but reasoning differently, Wineburg (1991) attempts to explain the reasons behind the difference. He discounts the idea that historians possessed and activated an appropriate principle-oriented knowledge structure suggesting that they did not use existing schema but carefully and thoughtfully constructed these tailored to the event under study. Differences amongst the two groups reflected different conceptions not of what a text *says* but what a text *is*. The key difference pointed out between the students and historians was the use of attribution as the question from which all else emanated rather than as an add-on as students appeared to do. For historians using the sourcing heuristic did not appear to be a rule or problem solving strategy but demonstrated a belief that texts are defined by their authors' Wineburg (1991, p.83). Maggioni et al. (2004) found teachers or pupils using methods in the absence of understanding the role or purpose of evidence.

The emphasis in Wineburg (1991, p.83-84) argument is thus the underpinning epistemic concept without which the procedures of the historian could not be applied meaningfully. The concept of what the text *is* was seen as the defining difference between novice and expert. This is an important distinction and can be said to support my argument that weak knowledge and skill in the use of heuristics alone may sometimes blur the underpinning awareness of the

importance, in principle, of evidence and argument in constructing an account. Student teachers, I found, did not develop the expert level of skill and knowledge in my workshops as indeed would have been difficult in the short interventions of three four-hour sessions. Many did, however, become aware of the use of evidence and argument as providing a means to judge between explanations. They quoted some heuristics and phrased responses in ways that indicated that a change had taken place from a Naive Subjectivist stance to one where use of method and heuristic was seen as a way out of relativistic helplessness.

Maggioni et al. (2009a, p.206) describe their examination of the stance of history professors expecting them to take a Criterialist epistemic stance. They found that historians acknowledged the need for criteria in the interpretation of the past but they rejected the view of history that equated interpretation with subjective opinion. Comparing novices to Maggioni et al. (2009b) historians in my data, I see that novices were more likely to be aware of bias and perspective at pre-test, more often not able to make this distinction between interpretation and subjective opinion.

2.6.6 Development and Validation of the CBKH, the Categories of Beliefs about Knowledge in History

2.6.6.1 Background for the Development of the CBKH

A set of open-ended questions described by Mason (2002, p.315) and used in early research by Bombi and Ajello (1988), as well as Boscolo and Mason (2001), were selected as measures of epistemic belief. These are 6 open-ended questions that are general, not specific context related, and relate well to a critical reader's experience of history. The questions are: What is history? How do people who write history know about the past they write about? What problems can historians have when they try to understand what happened in the past? Is it possible to explain what happened in different ways? Why? If there are two

different explanations, how is it possible to understand which is better? These questions require a qualitative analysis and can also yield scores on a continuum.

A rubric to differentiate between responses to the Boscolo & Mason questions was required. As the number of participants after all data cleaning was 83 and the sum of response sheets pre-test to post-test was 166, it was necessary that a format that could be used as consistently as possible be developed. The questions are open-ended and employed in a context of teachers and school history teaching, the indicators for categories needed to be specific where possible.

In order to assess responses, I had originally planned to use the framework devised by Kuhn and Weinstock (2002, p.121). Later, I studied Maggioni et al. (2004), and Maggioni et al. (2009a) model along with the possibility of transitions between stages as suggested by them. Based on this work and especially the study of Maggioni for her PhD as well as online discussions with her, I adapted their three categories and transitional stages to seven with definitions for the transitional stages. This development is described in detail in the Chapter and online discussions are cited.

2.6.6.2 Whose Beliefs?

A question raised by Maggioni[33] in discussion was on the CBKH: is this scale underpinned by the beliefs of the historian writer or the reader of history? Technically, this scale is created to access the beliefs of teachers as readers or users of history accounts in this research. The Boscolo & Mason questions the teachers respond to in the instrument, refer to a historian's problems in writing history accounts in Questions 2, 3 and the judgment between different explanations in history in Questions 4, 5, and 6. The reader's awareness of the historian in the text and indirectly, the historian's awareness of audience and purpose are discussed as required in the context of the argument.

[33] Maggioni, L. (2013). CBKH (personal communication)

The subjects of assessment could be teachers of history as are the participants of this research or their students or other readers of history. In practical terms, as I understand it, the historian writer of history texts, or even the pupil constructing an account in class using evidence and source would need to hold sophisticated beliefs in order to produce sophisticated accounts that may be considered credible by sophisticated readers of history. Without the relevant beliefs and skill in the production of accounts underpinning the practice, the accounts produced would have many underlying weaknesses. In judging a history account too, a naive reader would differ from one who was a sophisticated reader in making an informed judgment.

The purpose the CBKH serves is to access the stance in the given sample, the state of this sample and not a hypothetical sample. This compares the various responses within the sample to judge between them. Having said that, as a rubric to assess epistemic beliefs, it is structured in a rationale and logical progression from a state of naive to sophisticated beliefs that can be argued to be in keeping with the theoretical positions found in the literature and described in research studies as prevailing in the world.

The words naive and sophisticated are used per force for want of a better alternative. Sophisticated is an antonym of naive and therefore, these can represent the two ends of a scale. I recognise the judgmental connotations but numbers or letters alone would be inadequate.

2.6.6.3 Developments from Maggioni et al Model to the CBKH

The original model developed by Maggioni et al has three categories: the Copier, Borrower, and Criterialist. This is described and discussed in the earlier Section of this Chapter. Maggioni et al. (2009a) had expanded these categories to include transitions in their research. These transitions made sense after my early attempts at data analysis. I first expanded the three categories to the following seven.

Copier; More Copier than Borrower; More Borrower than Copier; Borrower; More Borrower than Criterialist; More Criterialist than Borrower; Criterialist.

This expanded model was used to assess the data but as I read and reread, recalled the sessions with field notes and the snippets of audio record, it became apparent that the model required more adjustment that could better differentiate between the responses. There were small adjustments required in the definition of the stages and in the spread from one level to the next. The transitions were there but the transition stage required more elaboration and definition. The following modifications were made to the original model.

First, I changed the titles of the categories. From Copier, Borrower and Criterialist Maggioni et al. (2009b), I changed to the straightforward: Objectivist, Subjectivist, and Criterialist in keeping with terms also used by Maggioni et al. (2009b, p.6-7). My readers in Pakistan may have difficulty in associating the same meanings to the terms Copier and Borrower that are intended by the authors as one needs to be familiar with the rather complex literature behind them. I reverted to more generally understood labels, Objectivist and Subjectivist, even though these do not capture the fine nuance of meaning the authors describe. The term Criterialist was retained being more self-explanatory and easier to defend. The terms Naive and Sophisticated are used, for want of better terms, to serve as easy access to two ends of a scale of development of epistemic beliefs.

Next, I considered what appeared to be the transition stage as more of a zig-zag between a naive and critical awareness rather than a smooth, linear, forward movement of development in belief. Maggioni and I discussed in mail[34] if the Borrower was more sophisticated than the Copier. In this sense the order and linear movement is problematized. This is discussed below and the zig-zag movement from naive to critical then naive again in the CBKH categories reflects this developing idea. In this order, an objectivist could be seen as naive to begin with but with the introduction to doubt in the naive realist stance could be called a critical objectivist as the individual has begun to doubt naive objectivist belief but

[34] Maggioni, L. (2013). CBKH (personal communication)

has not yet acquired the complete awareness of the constructed nature of knowledge that defines the subjectivist stance.

To reason the order, I argue that when the individual becomes aware of the mediation of the historian in the construction of knowledge in history, aware of perspective or position, or standpoint in the evaluation of evidence, the use of the methods of the historian acquire meaning and purpose. At the top end of the scale, the Criterialist is not just a more expert Objectivist but one who integrates aspects of the knower, known, with way of knowing as is argued by Kuhn and Weinstock (2002). It is with this reasoning, I retained the development as proceeding from an Objectivist to Subjectivist to Criterialist stance as a proposition that the interim phase of awareness of the knower is necessary for the individual to recognize the purpose of evidence and argument in the construction of history. It is not necessary to see the Objectivist or Subjectivist as more sophisticated than the other; they can be conceived as different but necessary, interim phases in a development of epistemic beliefs. Both the Objectivist and Subjectivist can be seen as naive, as Maggioni points out, but moving towards sophistication with integration of the knower and known.

The unraveling of the ‘way of knowing’ from concepts of the ‘known’ and the ‘knower’ in the assessment of development of epistemic beliefs about knowledge in history is a key change argued in the CBKH. This is discussed in detail in the response to Question One, ‘Why is it necessary to unravel elements of belief in knowledge in history?’

Spread in the stages of development in the Criterialist stance became necessary as it became apparent that participants who had moved on from an Objectivist, Subjectivist stance could not yet be called Criterialists in the full definition of the term as they had not yet achieved knowledge and skill in the methods of history and the heuristics that are employed to evaluate evidence and accounts in history. This is discussed in Question Two, ‘Is the Criterialist just a more expert Objectivist?’

Conversely, I was unsure if a set of response that displayed good knowledge and skill in the use of heuristics was necessarily a Criterialist if the

responses contained no reference to the constructed/interpreted nature of the account. Was it possible that the respondent, having learnt good use of method, had reverted to a belief in an objective account where the mediation of the author was completely overcome? Maggioni et al. (2009b, p.11) point out that the Criterialist would recognize the interpretive role of the historian. Awareness of the need of the acknowledged presence of the historian in the text is again argued below as higher level of sophistication that needs to be reflected in the category scale.

2.6.7 Question 1: Why is it Necessary to Unravel Elements of Belief in Knowledge in History?

Maggioni et al. (2009a) conceptualise development in epistemic beliefs as integration of the objective and subjective dimensions of knowing. This they express as the integration of the knower and the known in their model, the Copier, Borrower and Criterialist. Maggioni et al. (2004). Awareness of the presence of the knower in the construction of the history account and the constructed nature of the account are the two key elements mentioned by the authors. There is, however, an underpinning element of *how* the account is constructed with understanding and skill that is considered key in this process. This process is described at length in their writings. I have felt it necessary to unravel the process of integration by examining beliefs about the knower, the known, and ways of knowing in history as separate and *not necessarily congruent* developments in concepts, knowledge, and skill.

Suggesting that levels of knowledge and skill can be temporarily unraveled from concepts of the knower and the known is not to suggest that the 'symbiotic' relationship does not exist between 'knowing that' and 'knowing how' as it is conceived in the domain. What I have done at a practical level for one, is to call 'ways of knowing' a separate element and place it in a separate row after the known, knower for closer examination. What I argue is, that ways of knowing differ in levels of knowledge and skill amongst novices and experts and this difference

can be assessed relative to the data in a wider spread that recognizes the epistemic development but gives concessions to the incongruent weak knowledge and skill. I consider that a weak level of knowledge and skill should not be seen as precluding a move from a Borrower to Criterialist stance. In examining a respondent's beliefs about knowledge in history, when there is seen evidence of the presence of the epistemic belief that knowledge in history becomes possible by or through the use of evidence and argument this makes it difficult to rate the examinee as *not Criterialist as the specific use of heuristics is only weakly known.*

I refer once more to Kuhn and Weinstock (2002) conceptualization of epistemological thinking as the coordination of the objective and subjective dimensions of knowing. The use of the words 'by' and 'through' suggests a possibility that such coordination, albeit tentative, has taken place. If the words are followed by a reference to authority as arbitrator, 'ask the teacher' as suggested by one respondent, that could be seen as naive in its antecedents. Seixas (1996) points out that the second basis for historical epistemology other than traces lies in expert historical authorities. But where the reference is to evidence or argument or to an adequate heuristic as a means to select between different explanations, the move from naive to a level of sophistication has been achieved. What should follow is growing expertise and with it confidence in a sound account.

It is important to remind readers here that the CBKH orders the criterialist category at three levels, Naive Criterialist, Critical Criterialist and Sophisticated Integrator. This, I argue, as necessary for development of several aspects of belief.

I realised upon reading responses of participants that practical knowledge about the way of knowing was weighing heavily in assessment of epistemic beliefs in the responses using a category definition that required sound knowledge and skill. There appeared to be conceptual awareness of the good use of evidence and argument, of the search for reliable evidence, in the construction of history accounts. What VanSledright (2002, p.111) refers to as procedures or strategies were however weak in my respondents. The heuristics themselves were not well known even at post-test. This was a dilemma as clearly many respondents had moved out of a purely subjectivist stance with relativistic perceptions of knowledge

to a sense that evidence and argument could make history accounts possible. Many used the words, reliable, authentic, in reference to source or evidence as a way of judging an explanation as better than another. They could not therefore, be placed at Level 3 of a Naive Subjectivist stance or even at the beginning of doubt at Level 4 where a person begins to recognize a way out of the bind.

Therefore, I altered the description of the category of the Criterialist to the purely epistemic awareness of the use of heuristics to use sources to find reliable evidence even though expert knowledge or skill in the use of the key heuristics involved was not yet observed. I assigned expertise in skill to a last category, the Sophisticated Integrator.

I set a criteria that if a response to the question, 'is it possible to decide which explanation is better', the response is yes and the following words are by, or through, and then suggests, use primary or secondary sources, the respondent has crossed the epistemic barrier of relativism into the next level of belief where a criterion is seen as useful to resolve the problem of subjective knowing. This becomes easier to understand when the definitions of stance are examined. A Copier is one who is not aware of the knower, of the constructed nature of historical knowledge, not aware that conflicting accounts could conflict not necessarily due to 'false evidence' but due to a different reading of the evidence from the standpoint of the historian. The Borrower at the next level is aware that history is an authored account and the positionality, perspective, act as filters in the construction of the account. This in a sense is a loss of innocence and can lead to infinite regress where no account can be trusted. The Criterialist at the third level is one who has come to recognize the use of evidence and argument to construct a useful account.

If in response to the question, 'how is it possible to tell which explanation is better,' the respondent says, '*by* using evidence, primary and secondary sources', lowering this response to the level of the Borrower or Subjectivist is not appropriate as the person is now aware that some process can be used to differentiate between accounts and can suggest some heuristics. Lowering him or her further to an Objectivist level is again inappropriate as the person is aware of the interpreted nature of knowledge, and has passed over from naive belief in objective

knowledge in history. The person is now aware of criteria making it possible to choose between conflicting accounts but does not yet possess expertise in this action. Epistemically, the person has moved on from a naive helplessness to some confidence in the possibility of historical accounts but may yet not know or understand the process of doing so at the level of an expert historian. Therefore, what I suggest and have done in assessing responses is to lower the bar slightly to include such weak responses into the Naive Criterialist category and spread the levels of Criterialist belief from Naive Criterialist, Critical Criterialist to Sophisticated Integrator. Thus, the expert has a higher bar and teachers, not being experts, can occupy a lower level whilst still being included in the Criterialist category.

An interesting point here is related to Maggioni et al. (2004, p.188) analysis of a relativist position. They write, 'In this case, the constructed nature of history is acknowledged, but the historical method is not deemed an effective tool to deal with problems of conflicting or missing evidence'. I extend this use of the word 'deem*'* to my argument. Where respondents use the words 'by' and 'through' referring to some heuristic, they do in fact imply that they *deem* this to be an effective tool to use to find a better explanation. Evidence in the responses may show that this judgment is based on limited knowledge and skill but the person has moved on from a complete relativistic hopelessness to consider criteria as a basis for deciding between accounts. Here then, the purely epistemic value is differentiated from the procedural. This is complex however, and it would need to be seen if the judgment by the respondent was a considered judgment.

VanSledright et al. (2006, p.216-217) argue that epistemological standpoints must be studied with respect to the community contexts in which they are employed. They describe the BLTHQ as a tool to measure the standpoints of history teachers. They say that the 21 Items of the measure are about evenly divided on a rules-criteria-warrants structure of the disciplinary community of history and a knowledge-acquisition criterion of the school history community. They hoped to see which criteria the teachers' positions cohere with.

I see this comparison as purpose based. Where novices are to be compared with experts, procedural and strategic knowledge may need to conform to the level

of the expert whereas when novice teachers are compared amongst other members of their community, arguably, standards of skill and knowledge would be adjusted accordingly.

In my sample, teachers are not experts. I want to make the distinction that while they appeared to indicate, post-intervention, that they have moved beyond a helpless relativism to recognize the use of evidence and argument to judge between accounts they may not be able to do so at an expert level. Wineburg (1991) argues that this is complex work and not easily accomplished by novices. I have found that there are differences amongst novices at Levels 4, 5, 6 in growing concepts as well as skills and these differences can be mapped by spreading definitions of the categories in ways that are more responsive to the data. I have constructed the CBKH on this assumption and the assessment of responses is described in these pages. I leave it to the reader and possible user of the category scale to see if it works for them.

Finally, there is a practical reason for my interest in expanding the scale. My research is experimental, focused on supporting a hypothesis on the basis of observations of change. I need to differentiate in the sample at a finer grain. If I were to set the bar on expert level of epistemic beliefs to assess my sample of teachers, I may be reducing nearly all participants to a single row at the low end of the scale. By expanding levels with transitions of finer difference, I can differentiate in ways that is theoretically arguable as well as being practically more useful. Besides, as a teacher, I can visualize my students standing by my table to argue with me that the assessment is not fair were I not able to differentiate between them.

2.6.8 Question 2: Is the Criterialist Just a More Efficient Objectivist?

If the focus in assessment of an individual's epistemic beliefs with the CBKH category scale emphasizes one element only, 'ways of knowing' (e.g. the use of heuristics and method, the importance of the question in evaluating sources) as the key element to assign the individual to the Criterialist stance, there is a danger

that the Criterialist may begin to be seen by developing learners to be, just a more efficient Objectivist. It is therefore, important to retain awareness of integration of all three elements, knower, known, way of knowing, to award a level of sophistication. The definition of sophistication needs to be based on an emphasis on integration of the objective and subjective dimensions as suggested by Kuhn and Weinstock (2002, p.123) rather than on any single element. Maggioni questions (in mail[35]) that, 'If a participant is aware of a method, isn't this an indication of integration?' 'What could be missing here though', she argues, 'is awareness of the limits of knowledge and, probably, there is a misuse, reduction of the historical methods, too'. I agree that when a method is referred to, a knower is indicated as a user of the process; however, the constructed nature of this knowledge may not always before grounded in this awareness nor taken into consideration when the account is evaluated. It would be useful to add this cautionary note and look for evidence in the responses to the continued consciousness of the interpretation and epistemological underpinning in a Criterialist stance even if expert use of method is seen.

Giving importance to the tools of the historian, VanSledright (2011, p.66) argues that 'historical accounts will vary but can still be legitimate if they measure up to judgments about what people agree to believe (criteria) constitute good accounts'. Within this criterion, I would like to stress the importance of the acknowledged presence of the historian in the text and the tentative, propositional nature of the account. This argument is in the context of novice teachers, as readers and users of history, developing epistemic beliefs; however, on the level of expert historians, who are expected to be sophisticated Criterialists, the question, 'is a Criterialist just a more efficient Objectivist', can also be asked. Amongst histories produced by expert historians there will be differences in the degree of accounts that appear to be objective, true, factual, continuous, untouched by human hands, and those that are tentative, propositional, open to criticism and change. As VanSledright et al. (2006, p.225) explains, history is an

[35] Maggioni, L. (2013). CBKH (personal communication)

interpretive reduction. ‘It is necessarily tentative, unstable, and less than the past.’ A Criterialist would, by definition, be aware of the nature of history accounts as tentative judgments constructed with knowledge and skill being the best possible in the circumstances of fragmentary, discontinuous evidence.

History texts used in schools expect a naive, uncritical reading. VanSledright (2002, p.4) cites Scott and Barthes (1968,1986) to argue that the accounts historians create often give the impression that there was no interpretive process used at all and that their words directly capture the reality they depict. This is a constantly occurring concern in the literature. As I assess responses that appear to show that teachers have moved towards Criterialism with references to method as a means to choose between explanations, I can see some responses no longer refer to interpretation or subjectivity and most do not mention the tentative nature of historical knowledge. I see these responses as less sophisticated than those that do retain a reference to the perspective or positionality of the author. I realize that the Criterialist stance can be spread into degrees of sophistication, taking into account a continued awareness of the interpreted, constructed, nature of the history account.

VanSledright and Afflerbach (2005, p.16) argue in the context of children’s learning that to move students to greater sophistication involves shifting them from binary distinctions towards observations that historical accounts embody perspectives and subtexts that reflect the authors historicized positions. This awareness may get lost in the process of development if the focus remains on procedure as indeed has been found to happen. Maggioni et al. (2007) refer to a professional development program that required teachers to write a historical narrative. Teachers were taught to include detailed references to sources so that ‘it may force teachers to explicitly acknowledge the role played by evidence in the construction of historical arguments and thus foster epistemic awareness.’ Maggioni et al. (2007, p.6, 9) report that reading historical narratives written by teachers, they found, conveyed an authorless view of history typical of a Copier stance.

2.6.9 Question 3: Is Awareness of the Acknowledged Presence of the Historian in the Text a Necessary Aspect of Sophistication for the Reader of History?

The experience of meeting teachers of history and listening to their discussions and stories of classroom experience of teaching history in the interventional workshops was enlightening. Teachers described the perceptions of pupils and the sometimes emotional response to perspectives read in the textbooks. I realized that what is referred to as the effect of the 'referential illusion' on pupil learning needed to take its place as a principle in the overt and acknowledged presence of the historian in school history texts.

Concerns about the effect of the referential illusion have long been argued in the literature. VanSledright (2002, p.4) notes that 'the accounts historians create often give the impression that there was no interpretive process used at all, that their words directly capture the reality they depict, as though they were there chronicling every word'. In a current article, VanSledright (2014) restates the importance of recognising the role of the historian. He argues here for supporting knowledge claims by making evidence transparent via citation and subsequent checking by peer review. This is an important clarification. Maggioni (in mail)[36] argues that an attitude of recognizing our own positionality will seldom be cultivated if the role of the historian is not recognized. My participants do not know these words but spent much time in my workshops discussing pupils' concerns about the history accounts they read as a part of their curriculum and assessment requirements. Pakistan's past in relation to Britain colours the expectations of readers when they encounter history texts written and published by writers from the UK. Whereas there may be a ratio of pupils from European and African backgrounds in most classrooms in the US, the classroom in Pakistan is largely from one cultural background, Pakistani, and the text is from the UK. This enhances awareness of the positionality of the historian amongst pupils and teachers. Their concerns and indeed those of numerous writers in the subject area

[36] Maggioni, L. (2013). CBKH. (personal communication)

were used to redesign the last level of the CBKH to include awareness of the acknowledged presence of the historian in history accounts.

Those who could recognize and be critical of the referential illusion, should be recognized as more sophisticated I considered. This aspect of *requiring* the acknowledged presence of the historian and the tentative nature of the history account is arguably a higher level of epistemic sophistication. This is a sophisticated argument, and readers may be made aware of this shortcoming in history texts through the use of the CBKH scale. Assessment serves a developmental purpose.

Unfortunately, as most history texts read by students in school are evidence, the historian or the evidence-based, propositional, tentative nature of the account is seldom plain to view. Indeed it is this concern that has led me to include an enhanced emphasis on the need for transparency, of the constructed and tentative nature of history accounts within the CBKH.

The CBKH as a development from the models discussed and used by Maggioni et al. (2009a, p.196) in their research has several differences of emphasis, additional criteria, some unraveling, and argument. I owe the authors my heartfelt thanks and especially to Liliana Maggioni. With this I conclude my thoughts on the development of the CBKH category scale.

The literature review for this research required a study of a range of domains. Conceptual change literature was studied and a possible gap in knowledge was identified. Kuhn's description of the process of change of paradigms in scientific communities was seen to be different from the orchestrated change processes that are expected to lead to change in concepts in everyday classrooms. An examination of the difference in conditions in the two social situations illuminated the lack of a perception of agency amongst participants that could be a factor in change. A hypothesis was drawn from this and the research was designed to explore how this factor could work in teacher education practice.

The review of literature also necessarily covers epistemic beliefs and research within this as the particular concepts selected for study. Issues of theory

in mixed methods research and quasi experimental design lay the groundwork for design decisions.

CHAPTER 3

THE RESEARCH METHODOLOGY

3.1 Introduction

To recall briefly, this research examines a relationship between a perception of agency and change in concepts in contexts of task based discourse in teacher education classrooms. Problems of learning as change in concepts are a source of concern in my practice and indeed in education generally. This research focuses on change in epistemic concepts of knowledge in history in contexts of activity and talk to explore a gap of a perception of agency to question, agree, disagree, to reflect upon own prior concepts within the process.

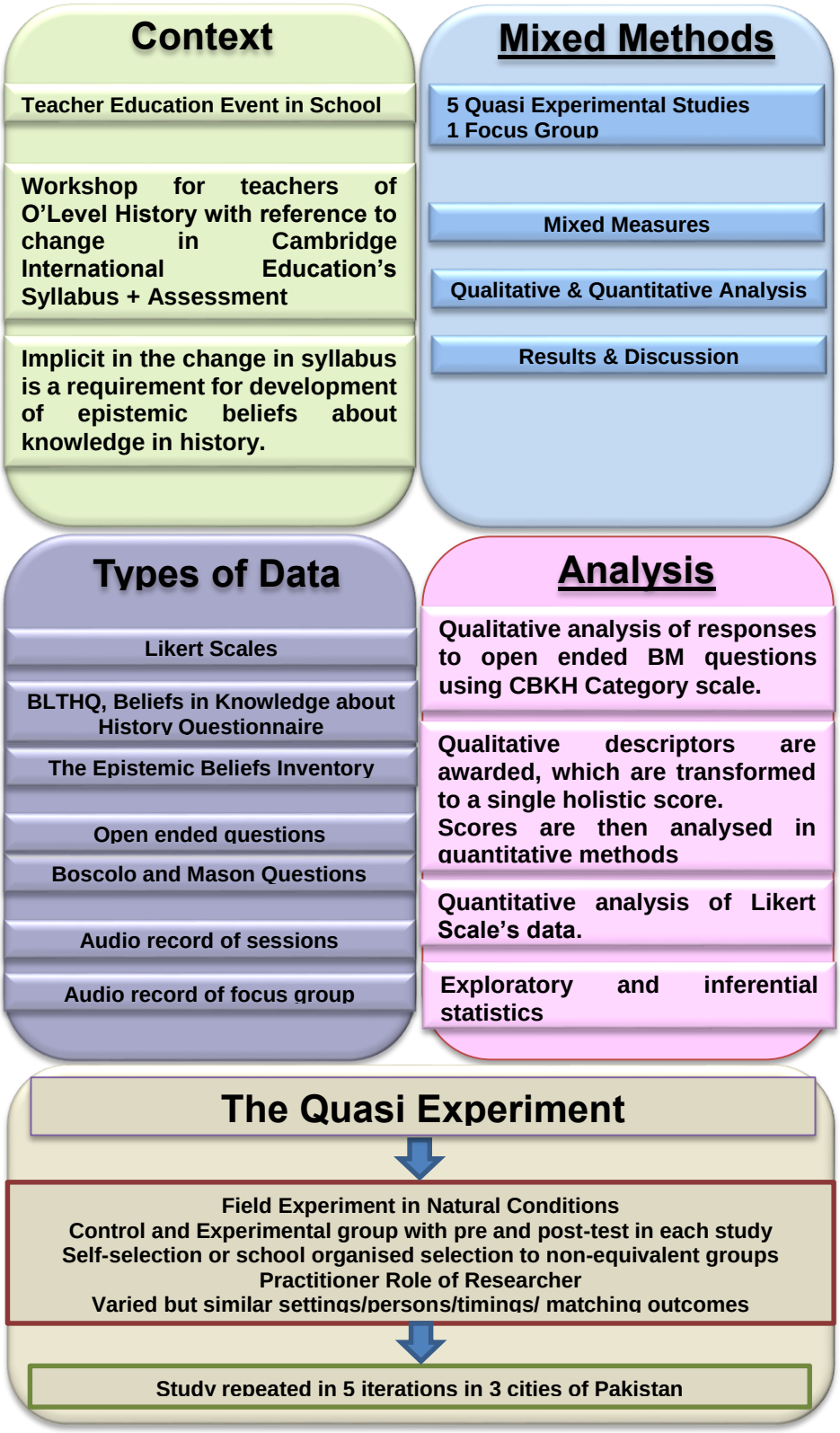
Task based discourse in classrooms provides the necessary contexts, actions, and words needed for concepts to change. Schön argues for reflection in action for convergence in meaning within contexts, actions and words. I base my argument for agency in task based discourse on a review of Kuhn's description of paradigm change.

A hypothesis is developed from the critical reading of the literature, and a quasi-experimental research to test the hypothesis is conducted and reported in this study. The practitioner research is designed to use real life situations for change in epistemic concepts. Workshops for teachers to prepare for syllabus change in history are the contexts. The research is designed as 5 quasi experimental studies which supports the drawing of conclusions and demonstrates how the relationship can be researched.

This Chapter on The Research Methodology describes and, where required, argues features of research design, the methods used, contexts of study including constraints and opportunities, measures and ways of data collection, sample and control, and aspects of reliability. Tables, a Map of Research, and Lists of features are provided for ready reference. The interventions and standardization is an important feature in the research and detail is provided in this Chapter. The

quasi experiments conducted are described in considerable detail as is necessary in this type of experimental design to aid in interpretation. Threats to interpretation in quasi experiments are necessarily discussed under headings as raised in the literature. The contexts are detailed with transparent descriptions of strengths and limitations. In the end, reflective conclusions are drawn with reference to the research methodology. Additionally, this Chapter presents diagrams of design features and tables to outline and display key features of the design, (see Figures 3.1, 3.2) the objectives, the interventions, and the validation exercises for assessment of reliability of qualitative assessment. This was seen as necessary as the research has a complicated design with five experimental studies with 5 Control and 5 Experimental groups, two interventions and a range of analysis. The data and analysis is both quantitative and qualitative. I request readers to first study the graphic representations for ease in understanding.

Figure 3.1: Features of the Research Design



3.2 Aim of Research

To enquire into a problem of teacher learning as change of concepts that I and my colleagues encounter as practitioners in teacher education.

3.3 Study Question

The main research question is if there is a relationship between a perception of agency and change in participants' epistemic beliefs in learning contexts of task based discourse.

3.4 The Research Hypothesis

The research hypothesis is stated as an alternative hypothesis in place of a null hypothesis. The hypothesis states that there is a relationship between a perception of agency and change in concepts in task based discourse.

3.5 Key Objectives of Research

- To investigate an alternative hypothesis of a relationship between a perception of agency and conceptual change in task-based discourse in 5 quasi-experimental studies conducted in teacher education contexts.
- To employ and examine Skemp (1971) principle of a range of suitable examples as a pedagogical method to change concepts.
- To demonstrate if the relationship of a perception of agency and conceptual change in task-based group discourse can be researched.

3.6 Specific Objectives for Conduct of Research and Analysis

- To conduct multiple (5) independent quasi experiments each with a Control and Experimental group, pre-test and post-tests and collect data.
- Use Skemp's proposal of 'experience of a range of suitable examples' Skemp (1971) as a common, pedagogical method to change concepts of knowledge in history as Intervention 1 in all 10 groups, Control and Experimental, in all 5 studies.
- Provide Intervention 2, an enhanced perception of agency, in 5 Experimental groups to test the hypothesis.
- Use Likert Scales EBI, BLTHQ, and Boscolo and Mason open ended questions as measures of epistemic beliefs and change within these at pre-test and post-test to collect data.
- Attempt an audio record of sessions and focus group to collect oral data to describe the process and participants' perceptions.

3.7 Objectives of Data Collection

- Obtain data to assess and describe the participants' epistemic beliefs and concepts of knowledge in history at pre-test and post-test and change within these.
- Obtain both quantitative and qualitative data.
- Obtain oral data of sessions and post study Focus Group to illustrate, if possible, awareness of a perception of agency from participants' perspective.

3.8 Objectives of Analysis

- Modify, develop, and employ a model of epistemological beliefs in history to analyse epistemic beliefs.

- Develop the model (CBKH) to assess and order epistemic beliefs as expressed in responses to open-ended BM questions.
- Qualitative analysis of responses to open ended Boscolo & Mason questions using the CBKH rubric.
- Assign quantitative scores to qualitative descriptors of epistemic beliefs as assessed using the CBKH scale.
- Carry out Inter-Rater agreement exercise and Internal Evaluation of assessment of responses to the CBKH scale.
- Analyse numerical scores through a range of statistical analysis to assess epistemic beliefs at outset and change within these, post intervention.
- Analyse data from responses to Likert Scales using quantitative analysis.
- Analyse data of all three instruments separately using a range of statistics for description and inference.
- Compare changes in epistemic beliefs from pre-test to post-test, in all 10 Control and Experimental groups due to Intervention 1.
- Compare differences in change in concepts between Control and Experimental Groups to study the effects of Intervention 2 and support the hypothesis.
- Compare differences in change between Control and Experimental Groups in 5 independent studies.
- Analyse change in general epistemic beliefs with the EBI data and epistemic beliefs about knowledge in history using the BM questions and BLTHQ from pre-test to post-test.

3.9 Features and Steps for Analysis

- Meta-analysis of overall combined data and independent analysis of all 5 studies.
- Comparison of change in stance independently in each measure.

- Quantitative analysis of Likert Scale data.
- Qualitative analysis of written responses to BM open ended questions to assess change in epistemic beliefs
- Transformation of qualitative descriptors into numerical scores for quantitative analysis.
- In-depth analysis of quantitative data through a range of numerical and statistical tests according to constraints of data.
- Summarize support for a possible relationship of a perception of agency on change of concepts using both quantitative and qualitative analysis of all data of change in epistemic concepts from pre-test to post-test.
- Explore attributions of change by participants as evident in audio record and focus group record in order to illustrate change.
- Analysis of responses to BM questions using the CBKH category scale. Descriptors awarded to each of 166 responses and transformed to quantitative scores for statistical analysis.

Note:

(1) The main question is drafted as sub-questions for analysis of data in order to draw conclusions. These sub-questions are provided in Section 4.3.2 in the Chapter on Results and Discussion.

(2) The audio record of the sessions is unfortunately not complete or continuous and often inaudible. Excerpts are taken and added.

3.10 Objectives for Discussion and Conclusion Based upon Qualitative and Quantitative Data and its Analysis

- Discuss difference in change in epistemic beliefs and concepts of knowledge in history as a consequence of a difference in a perception of agency in Control and Experimental Groups.
- Discuss change in concepts and relationship of change with a perception of agency.

- Discuss change using a range of suitable examples as a pedagogical method for change of concepts.
- Critically discuss the validity of measures and methods used in the study and the reliability of statistical conclusions.
- Discuss a perception of agency in task based group discourse as an educational intervention.
- Discuss the difficulties and possibilities of research into epistemic concepts and a relationship with perceptions of agency.

3.11 Research Facts

3.11.1 Design of Quasi Experiment

A quasi experimental, pre-test-post-test non-equivalent group design conducted in a series of 5 studies with both Control and Experimental groups in each study Cohen et al. (2007).

Within the educational context, which is a workshop, X marks the intervention, an enhanced perception of agency. This can be represented as follows for the 5 studies as in Cohen et al. (2007).

Experimental	**O1 X O2**	**O1 X O2**	**O1 X O2**
	----------	**----------**	**----------**
Control	**O3 O4**	**O3 O4**	**O3 O4**
Experimental	**O1 X O2**	**O1 X O2**	
	----------	**----------**	
Control	**O3 O4**	**O3 O4**	

3.11.2 Sample Size

83 after data cleaning, missing data is described.

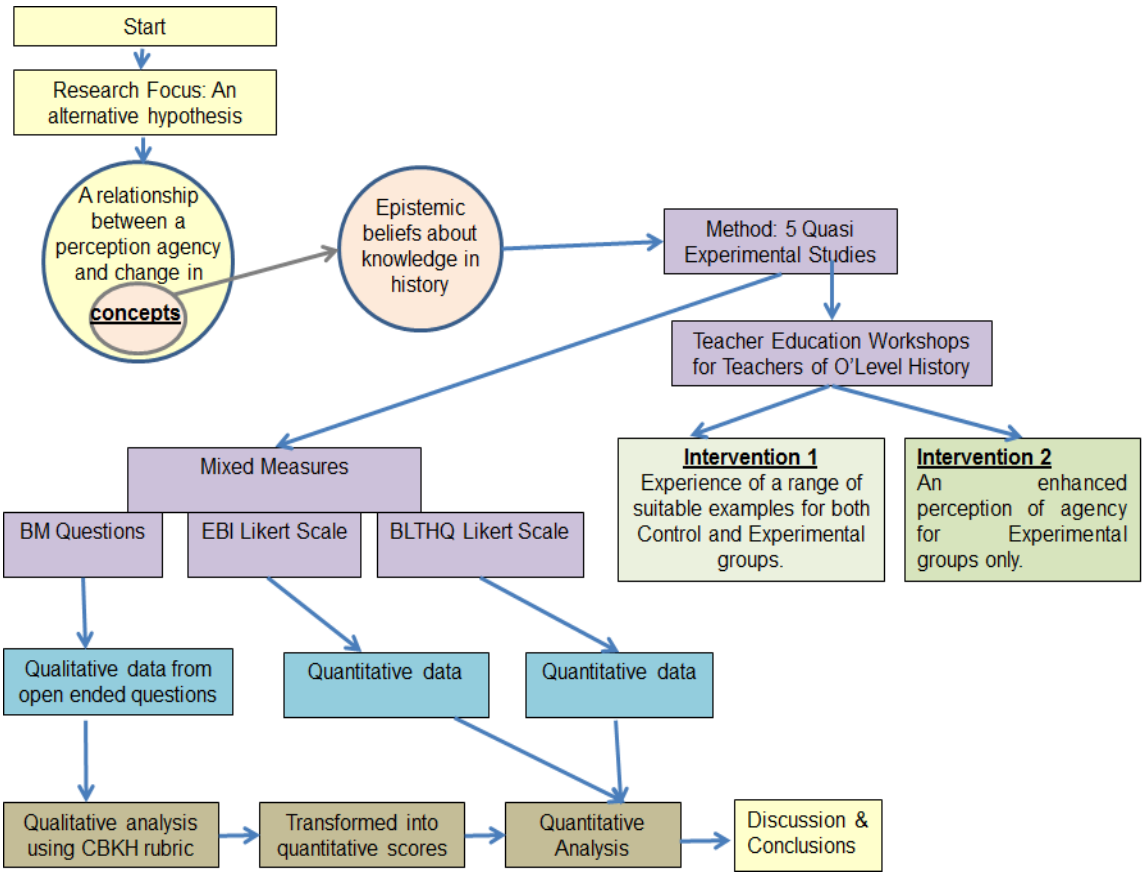
3.11.3 Dependent Variable

Change in concepts about knowledge in history: Concept of knowledge in history as isomorphic with 'the past' known and certain or a relativistic view of history as 'not possible to know' to an awareness of criteria and use of heuristics to examine claims and decide upon a better explanation.

3.11.4 Independent Variable

An enhanced perception of agency introduced in the Experimental groups only.

Figure 3.2: Map of Research



3.12 Features of Mixing of Methods Used in this Research

Mixed methods within each single study (repeated in a programme of 5 iterations of the same study with independent, different but similar groups of participants). Concurrent mixing at stages of design, measures, analysis, and drawing of conclusions. As a quasi-experimental study using mixed methods, there was a mixing of theoretical frameworks by design followed by the use of quantitative and qualitative measures for data collection and types of analysis.

The purpose of mixing methods was to measure the complex phenomenon of epistemic beliefs and obtain data that may emerge on a perception of agency. Triangulation and expansion of findings was made possible in discussion.

Level of mixing: At level of design, measures, analysis and at level of discussion. To begin, quantitative and qualitative data are analysed to their own logic and methods. Data is mixed first in transformation from qualitative descriptors to quantitative scores, which are analysed with statistical tests. Finally, mixing is essentially at the level of conclusion and discussion, where evidence from both types of data is used to support the hypothesis.

3.12.1 Quantitative Data

- Demographic data: Gender, age, years in service, educational qualifications.
- Responses to items on a Likert Scale on instrument, the Epistemic Beliefs Inventory, EBI. Authors: Schraw et al. (2002). (General epistemic beliefs).
- Responses to items on a Likert Scale in the instrument, the BLTHQ Maggioni et al. (2009a). (Epistemic beliefs specific to the domain of history incorporating aspects of pedagogical practice).

3.12.2 Qualitative Data

- Responses to open ended BM questions in a written framework as data to support the hypothesis Boscolo and Mason (2001) in Mason (2002, p.328).
- Transcript of Audio record of a Focus Group/feedback session with 5 respondents from an Experimental Group.
- Transcript of sections of audio record of sessions for first 3 sets.
- Field notes.
- Observation by a colleague for one day both groups only. The written data was unfortunately not submitted by the observer.

Note: Demographic data and the audio record of sessions and focus group are used to inform the write up and add quotations to illustrate the perception of agency and participants' learning.

3.13 Description of Research Methods

Research methods used in this study are described in this Chapter with particular reference to the constraints and opportunities that played a role in the evolution of the design and key features of method. The context of the research is described followed by a description of constraints and opportunities. The design is explained and the means of data collection are described in some detail. The intervention is described next with a summary of procedures to standardize this over the groups. Materials used in the workshops are listed and finally the key features of the sample and control are explained with affordances provided by field conditions as a rationale. Analysis will be described in the next Chapter but a brief summary of research facts does lay out the plan.

3.14 Answering Research Questions Using Mixed Methods

The responses to open ended questions about epistemic beliefs are vital data and very time consuming to interpret. The descriptors awarded to each of the 166 responses are transformed into scores and therefore, lost to view and qualitative methods appear to have less of a role to play in this research. However, in interpreting results, all evidence is employed. In this research, data from both measures corroborates findings and expands what can be said about the quality of change. Triangulation of findings is interesting. The instruments differ in that the BM questions are theoretical and abstract whilst the BLTHQ Likert Scale is set in contexts of pedagogical practice in history teaching. The EBI examines general epistemic beliefs. Therefore, change in responses are analysed independently in each instrument and then findings are compared with reference to change.

Greene (2007) argues that the decision that is most important is the level of interaction between the strands. Transforming qualitative descriptors to quantitative scores for analysis is one level of interaction, the other is in the conclusions drawn from the wealth of data in my research. Status or priority of different methods in weighting different results to answer study questions is important as suggested by Creswell and Plano Clark (2011). In the last stage of drawing conclusions from both types of instruments, the dual nature of the research exercise becomes apparent as interpretation takes over.

Here data of the two types of instruments and the different types of analysis are considered in tandem. The measures play a key role as they assess related but different aspects of the phenomenon as well as what could broadly come under the umbrella of general epistemic beliefs. In one sense, priority is given to the qualitative data as the responses were elaborated to some extent and appear considered and coherent. The data from the BLTHQ Likert Scale is related to the context of history teaching, but recognised as a forced choice response, with some social desirability issues in its pedagogical context perhaps causing dual positions to be taken by respondents. Thus, the status, if any, accorded is more on the basis of practical issues rather than a paradigmatic stance. It is a matter of belief in the

validity of a finding from qualitative *or* quantitative data that is a complex question to answer. A pragmatic stance is taken to interpret and draw conclusions from both sets of data.

3.15 Background

I, Lubna Kidwai, am the researcher. I am a teacher educator, and this practitioner research was designed in response to an often observed problem in the field. Over the years of our practice, my colleagues and I had noted that concepts are difficult to change and there was a need to make teacher education classroom experience more effective in change management. This problem of conceptual change was, in fact, the basis of my original proposal to the University of Bradford and admission for an MPhil leading to a PhD. Research into mechanisms that drive development in epistemic cognition is recommended by Pintrich (2002).

My prior experiences of the problems of conceptual change as well as extensive reviews of literature in the course of my study, led me to understand and formulate a relationship that could be studied for this research. I had observed, as the video[37] I present exemplifies, that social constructivist learning situations have dynamics that may negatively impact learning. Could a perception of agency relate to conceptual change? This was the question that came to mind. Reading in my years of teacher education and working on a Social Studies Curriculum had honed my understanding of difficulties of changing concepts that underpinned much learning in the domain. Kuhn's arguments for paradigm change, although not focused on change at an individual level, and certainly not on classroom learning or teacher learning, did help identify a relationship that could explain the difficulties in change in concepts through classroom learning. The problem was: Could this relationship be studied in empirical terms? I decided to try to test the relationship for my research.

[37] http://tinyurl.com/mdrpjvd

An important challenge was to find a concept that needed to be changed in the ambit of teacher education and to research this in ways that could facilitate the observation of a relationship of a perception of agency in the process. Concepts are hard to define and conceptual change is hard to put into a petri dish to observe. Essentially, schools are places for change and development of concepts but to observe a change relating to intangible phenomenon and control other influences is difficult. It was my good fortune to come upon a change in the history syllabus that was underpinned by an expected change in beliefs. How I managed to research this is described in this Chapter. Figure 3.1 and 3.2 can assist in understanding features of the design and the provision of the two interventions.

In the design phase, I considered a choice of identifying a context where change was expected and doing an observational case study. There were possibilities however, that change may or may not take place and factors other than a perception of agency play a larger role in the change process. The underpinning epistemic quality of the change in the syllabus being unknown to the teachers and schools, the chances were that it may not happen unsupported.

Another key factor in this decision was that I was researching conceptual change in my own practice, which is implemented in teacher education classrooms and not freely occurring social change. This led to the selection of practitioner research in a classroom setting.

An exploratory, quasi-experimental research into the issue was developed and mixed methods for data gathering and analysis were selected. This had necessarily to be in my own practice as it was difficult to find teacher education programmes and educators willing to permit access for research. Teacher education in Pakistan takes place either through government run colleges or through private non-governmental bodies. Unfortunately, most of these private organisations are either not run with professionally trained teacher educators, or run extended courses that could be the context for research.

The design itself was constructed to use a precious, small window of opportunity that arose with a change in syllabus and the resulting need for teacher development to meet the challenge of changing concepts. This provided a point in

time when a key, definable, conceptual change was needed in an accessible population; interventions to facilitate the change were possible and desirable; no other interventions that could possibly be threats to inference were on the horizon; instruments validated in previous research in the field were available; I, the researcher, had some access and ability to provide the intervention. Therefore, the opportunity was grasped and the research designed accordingly. As this is a completely self-funded PhD, a longitudinal, observational study of change in concepts as they unfold was also not feasible.

I chose a mixed methods research using methods and measures with which I was comfortable. Practitioner research is a tradition in schools and something that I had practiced and taught for a long time, therefore it was the best choice in the circumstances.

3.16 Locating the Study

'.....the sources are there, and we make questions out of them and sometimes we asked (pupils) what do you see in them and they make their own bank of questions from there.'

- Participant in focus group describing what they do with the extracts from sources that are now added to history text books in line with syllabus change.

In order to research conceptual change, I had to plan a training programme, and interest schools and teachers in it as I was refused access to research an existing course. For teachers and schools to be interested in a course offered by an independent researcher, a strong rationale would be needed and that is exactly what the change in syllabus offered. It is necessary here to explain why this was a time and space bound window of opportunity as many decisions were taken such as the selection of ready measures, a practitioner researcher role for myself in order to use the opportune moment.

Cambridge International Examinations (CIE) conduct O' and A' Level Examinations, for which private schools in Pakistan prepare pupils. The Pakistan

History and Culture Syllabus is being revised gradually to introduce understanding of multiple perspectives and interpretation and the use of source and evidence in learning history. This requires a shift from a traditional perspective of knowledge in history as isomorphic with the past to one of knowledge in history as constructed. Change in syllabus is not at present reflected in change in examination and assessment in significant ways giving teachers and schools time to prepare. This change in syllabus necessitates teacher learning of new concepts i.e. perspectives on historical knowledge. While the CIE had carried out some introductory training, currently there were no other programmes on offer in Pakistan in this particular focus. With time, schools and CIE would possibly be arranging workshops to meet the need for preparing teachers.

I recognized that training events such as those organized by examining bodies would have effects on the population that could be difficult to isolate. With more alternative courses on offer in this area, threats to valid inference would increase. There was no time to waste and the opportunity was taken but at some cost to the design. Written instruments to measure epistemic beliefs have been used in the world for some time and there was no published reference of their use in Pakistan. As this changing situation of teacher development in historical thinking did not allow the luxury of developing and validating new instruments in time, those available were put to use after initial piloting.

3.17 Procedures

The sample was approached through the offices of the school system. A poster (Appendix-A2) was sent for display to invite history teachers to attend History Workshops and schools were requested to make arrangements in their training rooms on the dates agreed upon. The poster informed participants of the research, explained cost and benefit, and gave details of dates, etcetera. The schools mailed lists of participants, who were divided into two groups by them more as a matter of convenience based on location. The history workshops were conducted over a three day period over three weeks for the first three sets; then

over three consecutive days in the last two sets. These were for 4 hours each. The workshops were held in the school's staff development centres, which are located in the cities of Karachi, Islamabad and Lahore. Teachers from school branches in the systems travel regularly to attend training from small towns in the periphery. As is the practice, travel costs were paid by schools and a meal depending on time, and tea was served. A total of 5 sets with two groups each, one Control and one Experimental, were held. Thus, there were 10 groups in all with 83 participants (after data cleaning). The second term of school, from January to May 2011, was the time frame to complete all 10 workshops, as these were called, for the 5 studies.

The pre-test was conducted as explained below on the first day and the post-test at the end of proceedings on the last day. The instruments were distributed in the training room itself and participants of the workshops were requested to fill these. As the facilitator, I remained present during the test, answered any question that was asked but generally stayed quiet. It took about 25 minutes to fill all three instruments, which were collected on the table. The post-test was recognized as being the same and participants often commented and sometimes asked outright why they needed to do this again with knowing smiles and occasional laughter.

In the Experimental Group, the Intervention, a perception of agency, described below was enhanced and sustained over the three days.

3.18 Research Constraints and Opportunities

The number of teacher educators accredited to deliver educational programmes in the private schools, to which I have access, is small and there were none with experience of history learning and teaching apart from myself. To be able to prepare other teacher educators was not seen as possible in a short time scale especially as the persons would need to hold the relevant beliefs, knowledge and skills of history. This was therefore, a forced choice. It was also seen as useful that the intervention could be carefully conducted with the researcher as

practitioner, maintaining the standardization of instructions, as well as, as much constancy in the teaching procedures and environment as possible.

Ready to use instruments, non-random, non-equivalent groups as found, varying field conditions, small sample size, make this less than a 'true experiment' limiting the kind of inferences that can be made. At the same time this has its advantages in the representative nature of the sample and the value in normal field conditions. Another advantage is the comparability of obtained data with that emerging from other research around the world. Valuable findings on existing beliefs and possible change in action can be found in published literature.

3.19 The Design and Data Collection (Appendix-A6)

The research was conducted as a quasi-experiment with a pre-test-post-test Control and Experimental group design Cohen et al. (2007, p.276) in a series of 5 studies with 2 groups i.e. a Control Group and an Experimental Group in each study carried out as a series of five small studies. The experiment measured a change in concepts in the setting of a learning experience, which was based on experiencing a range of suitable examples of different explanations of events in history with the use of sources. A perception of agency is the independent variable provided to the Experimental Groups alone. This learning workshop was provided to all participants in both groups, Experimental and Control. The Control Groups however, did not receive the intervention of an enhanced perception of agency. The design was developed with constant reference to Cohen et al. (2007), Cook and Campbell (1979) as well as Shadish et al. (2002). See Figures 3.1 and 3.2 to understand key features.

The pre-test and post-test was set for both Control and Experimental Groups. The test measured concepts of knowledge in history and the items could raise awareness of this concept even in the process of measurement. Therefore, it was necessary for both groups to experience this at the outset. It was not necessary to hold back the pre-test from the Experimental Group as the test's content made no reference to the independent variable, the 'perception of agency'.

The test could not affect the group's perception of agency, which was the independent variable apart from a possibly perceived lack of choice in actually doing the questionnaires or completing all items that may be challenging. It is for this reason, the pre-test questionnaires were mentioned in the initial poster and handed out at the beginning before the intervention procedures began. The procedures (detailed below) to enhance perception (for the Experimental Group only) were conducted after a relaxed 5 minute session of talk and humour to change the subject.

Educational research to study the value of interventions is a common enough practice as is assessment of learning. Therefore, the choice of pen and paper measures, and their pre-test and post-test use, did seem very much in character in the world of schools. Teachers looked quizzical at first, made a few jokes but filled them in with some interest, occasionally with some comments amongst themselves. The pre-test was conducted as soon as participants settled down and the workshop was begun and the post-test was held on the third day at the end of the day. I am happy to report that all participants filled the questionnaires and many responses showed careful reading and were quite reflective. The neutral option was not overused and no more than one or two response sheets had a single option marked on items of the Likert Scale. Nowhere in the responses did it appear that the question or statement had not been understood although words such as facts could possibly be analysed as to the meaning made of them.

To measure concepts about history, 'The Beliefs in History Teaching and Learning Questionnaire', (BLTHQ) Maggioni et al. (2009a), was used. Triangulation was achieved with the addition of an instrument to test general epistemic beliefs, 'The Epistemic Beliefs Inventory' Schraw et al. (2002), as well as a most useful set of open ended questions for a more deliberative though structured response. These questions have been used as part of research and are referred to by Mason (2002, p.328).

It was also a stated intention to use performance assessment in tasks to enhance the data. This is a common data gathering device in the domain and provides a useful comparison of reported belief with in-task application. This could

not, however, prove useful when conducted and was dropped as three day workshops did not provide enough time for participants to learn the use of the heuristics involved in the task competently. I would, however, recommend these to other researchers.

The workshops were constrained to be attractive to teachers and schools, not just for them to sign up but to continue to attend all three days. To develop concepts in depth and skill in the use of heuristics, however needed an exclusive focus on source work. Courses that VanSledright and others describe to prepare teachers for history teaching are university based and longer term. The focus on source and evidence in history was interesting for teachers but not a key part of their curriculum, therefore it was necessary to add tasks related to examiner reports, etcetera, which is time consuming. To resolve the quandary, I kept the content and objectives ostensibly focused on the kind of history teaching and learning concerns that make sense to teachers rather than a wholesale focus on source work that they did not seem to think important enough. The poster sent to schools inviting participation reflected this concern. Source work occupied a third of workshop time. In the remaining time, activities such as moderation of marking and assessment issues, were included to keep the workshop meaningful. At all times, a worry was 'will they return on the next two days of the workshop to be present for the post-test?' Luckily, after the first study, nearly all those who began did see it through, finding the content interesting and challenging. The concepts did develop as is evident in all measures, but the shortage of time limited the development in knowledge and skill in evaluating sources with appropriate heuristics.

3.20 The Intervention and Standardization (Appendix-A6)

The main intervention under study was a carefully enhanced perception of agency for the Experimental Group only. It is necessary to emphasise that a perception of agency was limited to an enhanced 'perception' in reference to an opportunity acceptable and useful, freedom in questioning and making sense,

opportunity to reflect upon own prior ideas and the new ideas discussed. It did not specifically imply for this study, physical arrangements, grouping, task differentiation, talk time, or other social aspects of classrooms. Participants were only made more aware of the condition of their intellectual agency in relation to the ideas being introduced.

In spite of much searching, I could not find existing instruments that could be used to measure a perception of agency during classroom learning in adults. Therefore, it was decided to measure the change in concepts in pre and post- test with the enhanced perception, an artifact embedded in the change intervention. This is described below.

A perception of agency is a psychological construct and intangible, therefore, it needed to be operationalised for the research. After introductions and the reading of objectives, and crucially, filling the instruments at the outset of the workshop, a statement was made that this is a workshop not a training session. This was explained as follows:

> A workshop is a place where people collaborate and work together in order to learn and construct meaning. Opportunities for talk are provided to better articulate and understand personal beliefs and practices. These workshops are contexts for meaningful tasks in collaboration and discussion. This is not a training session or lecture where knowledge is delivered.

3.20.1 During the Session the Following Steps were Taken

i. Participants were encouraged to ask questions, to agree or disagree with group views.
ii. Participants were reminded during group discussions to think and reflect how an idea related to their own prior knowledge, and try to make links with it.
iii. Participants were encouraged to raise their own examples.
iv. Participants were encouraged to consider if particular content was personally useful and meaningful to them.

v. Participants were reminded that they could slow down and pace their work to understand what was discussed.

vi. Seating was in freely formed groups and my own role remained that of a facilitator. Tasks were carried out in small groups to examine sources and conflict in history accounts. I led whole group discussion but I avoided developing a single whole class account of an event.

vii. As the facilitator, I deliberately paused at important points and waited to allow participants to decide rather than impose a point of view.

viii. Participants were reminded in tasks that they can change groups and work with people of their own choice. At the beginning of each of three workshop days, I kept circle time and then encouraged them to form groups of 4 or 5 people and move around. The seating was in small groups though different grouping arrangements were encouraged according to tasks. I observed that not many changed groups, perhaps not to offend others or because they were comfortable but the concept had been introduced.

ix. Participants, who were in whole group discussions and small group tasks, were repeatedly encouraged to disagree/agree with the facilitator and other members of the group rather than passively accepting what was said. A wonderful snippet from my field notes illustrates how I could see this working. One small sized young female teacher, raised her hand high in the middle of a whole group discussion and announced to the whole class, ‘I disagree!’ When I focused on her and asked her to elaborate, she went ahead and made her point.

Note:

An important point to note here is that I as researcher had a very limited access to schools and was dependent on school administration to contact teachers and make arrangements. While schools were sent the poster and told in writing that the workshops needed to be voluntary, it was difficult to know how voluntary this was in real terms. It can be suspected that schools arrangements differed in, informing/inviting/persuading/even charging teachers to attend the offered workshops. On the other hand, as the word

got around regarding the workshops, there were in most cases more attendees on the second and third days than the first making a case for personal motivation. As there is a dearth of professional courses for teachers of senior schools, it is possible that the workshops were generally welcomed. A record of the last feedback session provides useful data in this area.

The first study, however, was an exception where several participants from the Control, and some from the Experimental group, did not return after the first day. The reason could possibly be that these were members from a large school branch with high rates of success in CIE examinations and they did not perceive the 4 mark change in Syllabus to include questions on evaluation of source in history as consequential. This was a point that had come out in the pilot I had conducted earlier. Importantly, the percentage of drop in members in the Control group (21.42 %) was lower than in the Experimental group (50%) in the first study. In the third study, the Control group lost 4 of 15 while the Experimental group lost only 1. In the remaining 4 studies attrition after workshops began was no more than 1 or 2 members of a group.

I consider that it was perceived that the workshops were less closely tied to examination content than earlier thought, which is why more participants dropped out in the first study. In this city and school system, performance in examination is high and teachers are highly prized, being paid large salaries. Those who dropped out in the first study are among the more successful and have a voice. I consider it a part of my success that so many attended at all. It must be noted however, that from the original estimate of possible attendees, only a third came. Security and cost implications apart from school constraints are also a reason people did not come from out-station schools.

I did not consciously change the content, pedagogical method, or intervention in any way from the first to last study. It is possible that I grew in confidence after the first round of two workshops. I think I was nervous in the first session, which I was not later. I have a long experience of teaching and teacher

education as well as the development of social studies curricula. I am generally a popular teacher, well known, which must have been a factor in the learning but common to all groups.

3.20.2 Experience in the Control Groups with Reference to the Experimental Treatment

The first Intervention, experience of a range of suitable examples was provided to all 10 Control and Experimental groups. Apart from this, workshop format, task-based group discourse, materials used, activities, content of talk, number of days of the workshop, myself as tutor researcher, bilingual language, were features maintained as common in all groups. The numbers of participants, building environment and comfort levels, pace of lesson, incidents, food quality, distance travelled, cultural, language background and personal characteristics, etcetera, had multiple random differences that did not have a pattern and were not possible to control or measure. Both Control and Experimental groups of each study were conducted in the same city. Islamabad-Rawalpindi are taken as one, being twin cities, sharing many cultural and geographic features. The first three studies were conducted over three weeks with one day a week. The last two were conducted on three consecutive days due to school constraints.

3.20.3 How was the Experience Different in the Control Group?

How the experience was different in the Control group is an important question. First of all, I kept my focus on whole and small groups without giving particular attention to encourage individuals to think or question prior concepts. I did not consciously announce, encourage, or persuade individual participants to consider their own prior concepts in reference to group ideas. I did not ask questions of individuals as ‘what do *you* think?’ ‘What has been *your* experience?’ as I did in the Experimental groups. I did not deliberately encourage individual participants to reflect upon their *own* prior concepts about the epistemic value of

the content in the history books although that was the focus in the whole group discussion and tasks. I did circulate and provide input to groups without focusing on individual participation in talk.

I avoided focus on seating, movement, participation in talk, or pace for individuals to comprehend. I avoided noticing where in small groups, individuals were left out of the conversation or tasks or where a member or two appeared to be quiet or dominating the activity or conversation. My effort was to not develop a perception of agency to consider and reflect upon their own meanings, to differ, to question, agree or disagree at will, amongst participants as I did in the Experimental group. I kept the video I have mentioned of a group task in an early years classroom as a model in my mind. In this video, I interpret learners missing a perception of agency and this was the scenario I avoided in the Experimental groups. To be honest, there must have been occasions when I forgot which group I was in but that was not a common feature as the pressure of research kept me generally focused. That model video has been placed online and can be seen at http://tinyurl.com/mdrpjvd. The video is not edited and blurred at onset. The important section is the activity taking place in the last group. The camera focuses on this.

3.21 Materials Used for Learning and Teaching

The first intervention was a pedagogical device used in all 10 groups, Control and Experimental. Using Skemp's argument for experience to change concepts, it was decided to use 'a range of suitable examples' to facilitate conceptual change, Skemp (1971). Task-based group discourse was the context for learning. Three topics were selected from the CIE syllabus and well researched; rich source material from a range of perspectives was found and employed. Types of sources were varied and criteria to analyse sources as evidence was used. The same content and method was used in all workshops.

3.22 Topics and Examples of Multiple Perspectives and Source Material Used in Workshops:

- The Bengal famine/holocaust of 1943-1944 with two perspectives, one laying responsibility on the Bengal Muslim League Ministry, the other on the British Government. Sources: O'Grada (2008), Mukerjee (2010), Wikipedia.
- The secession of East Pakistan from Pakistan, 1971, with two perspectives on the causes of secession. Sources: Bose (2005).
- The annexation/accession of Kashmir to India, 1947 and the role of the British Government. Sources: Jinnah Papers with source documents Lamb (1992).

It is important to emphasize here the importance of the material selected. The examples were particularly focused and sharply indicative of conflict in sources and different explanations. The examples demonstrated how Pakistan History particularly, has been represented and interpreted in various ways, and how sources can be evaluated and used to construct and analyze an explanation. All participants read the material, evaluated the sources and drew conclusions which were discussed. The discussions took place first in small groups and then with the whole class pitching in. VanSledright (2002, p.140) suggests that to pursue learning by investigation involves choosing topics and events that hone powerful investigative, domain-specific strategies and then use these to build thought and critical reading capacities. I selected topics for the presence of conflicting accounts, availability of evidence, and relevance to the curriculum. These examples also have some emotive value as they relate to particularly painful events in the Sub-Continent's History.

3.23 The Sample and Control

In implementing the design, it was found that the construction of groups was not in my control. School concerns and commitments dictated the when, where

and with whom the research could be conducted. A range of common and varying features were found in the sample.

To study the influence of an educational intervention on conceptual change, the aim was to target as many members as could possibly be available for the three day courses. As Cohen et al. (2007), cite Gorard (2003), one can start from a minimum number of cases per cell to calculate the size of the sample. As this is an entirely new area of research in Pakistan, it was difficult to predict the size of difference in relationships. I feared that there will of course be non-response, attrition and respondent mortality and this did happen as is described. Some participants did fail to attend, or return questionnaires.

At the time of research planning, I did have access to the entire population of over 200 in-service teachers of O' Level History, in a single large school system with branches spread over three geographical regions in Pakistan. I was a tutor for University teacher education courses in the large Pakistan School System but a sudden rise in costs brought these to an end. Therefore, I lost direct access to the teachers in the school. The worsening security situation in the country also reduced the number of the teachers in the originally identified sample of 200 that could travel from out station schools to attend the planned intervention workshops in three main cities. This caused a setback to the research making the numbers low for statistics. The smallness of the sample made it necessary to reach out to two other similar O' Level teaching school systems and re-run the workshops to add numbers. In the end, 83 attended all three days and filled pre-test and post-test instruments.

There were broad features common to all groups. All in the sample were in-service teachers of history for senior school ranging from Classes 9 to 10 in the first 3 sets, and some from Classes 8 to 10 in the last two sets. These were from 3 large schools where the CIE Syllabus is taught and learnt by both boys and girls in English. Nearly all taught a section of Pakistan History although some who were found to be teaching to Class 8 taught a lower level. All were from School A in the first three sets and used mainly the same 2 history texts by two authors and had become used to seeing sources referred to in the texts. Some members in the last

two sets (Schools B & C, Class 8) used a different text but which also contained reference to sources. All schools were preparing pupils for the CIE examination. Some teachers in one group were seen to be teaching to Matriculation and not O' Level. All three schools had a commitment to professional development and some form of collaboration with universities in the UK thus an ethos of professional action could be assumed. IELTS scores tested for teachers, as seen on Courses, is an average of 6 although there would be a range.

Stated workshop objectives/broad sequence/group seating/tasks/task materials/arrangements of distribution of materials/presentation materials/ facilitator were common for all groups. There were differences in settings as all 10 workshops were held in different locations as organized by schools. These were located in three cities of Pakistan and therefore, there were cultural differences as well as first language differences. Educational qualifications differed within a band of history, Pakistan Studies, within a component of history or International Relations with a component of history. There were more female than male teachers.

Random allocation to groups was not possible nor was it possible to match groups. Most came with a Graduate or Masters in Pakistan Studies, International Relations, or Political Science and volunteered that there had been components of history in their courses. Two had studied Philosophy as a component and this is evident in the analysis. This information was not available from schools in advance and was obtained upon inquiry in the introduction sessions. Generally, participants did not appear to like volunteering this information except for those with relevant backgrounds. As the data spoken out loud, was a source of potential anxiety, I kept questions off hand and at low key, and written recording of findings could not be made. This could be a serious threat but for the knowledge that historiography is possibly not more than a part of the knowledge content and seldom taught as a skill or a development in epistemic cognition in University history courses for Masters degrees anyway. No teacher appeared to have a moderate or high level of skill or understanding of heuristics at entry. Some had theoretical understanding. One who had studied history as a subject referred to

Durkhiem, and another wrote a phrase in the post-test that suggested he/she had been reflecting and reading up on epistemic aspects of historiography.

There was a range of differences in Universities or Colleges attended but this was not questioned. There was a range of differences in the cities and towns they came from and the distance travelled to attend. I insisted that workshops for the two groups (Control and Experimental) in each set were either on the following day or close in time and the time sequence of workshop sessions were the same between two groups in each set. *Schools were requested to form a set of two groups in each city and the decision was made that the first of the two in a set would be the Experimental group. This was adhered to in all sets.*

This was very much an experiment in the field and all kinds of unforeseen occurrences created a range of different situations. In the first set, a teacher, who was a part time employee with long experience and a pedantic style, focused on precise marks schemes, kept interrupting the session with arguments and demonstrating her boredom with the source work. Two attended the first sessions of the two groups in the first set saying they were not teachers of history but wanted to learn. They did not return on subsequent days and data from them is not added to the data set.

There were also two other participants travelling for nearly three hours who came all three days to attend the workshop who had also never taught history but wanted to learn to do so having taken Pakistan Studies at college. There was a girl from an out station school who drove over bad roads and a poor security situation to attend all days even though she had an accident one day. Their motivation was awe inspiring but their knowledge of the history content was limited. All three were from a Control group.

It is important to note that both groups, the Control and Experimental, attended the workshops. Both were given pre-test and post-tests. The content of the workshops, the sequence of presentation, the tasks carried out, the examples of historical source and evidence used, were kept as carefully matched as possible. I, as the facilitator, was common in all. No members of the administration were part of the participants or present in the class.

As the saying goes, one can never cross the same river twice. It is recognized that no two sessions could ever be the same. No two people could be said to have had the same experience. In many significant ways however, the facilitator being the same, keeping course content and material matched, it could be said as far as educational experiences go, they did all attend the same course.

How far participants perceived themselves to be agentic was not measured nor indeed how far the message of agency had been registered by each individual. With real life field conditions reducing the degree of control, the smallness of the samples and the multiplicity of variables making strong inference difficult, it was understood that a good quality description using both statistics and qualitative analysis may be the best that could be attained. A focus group session was organized and held with the last Experimental group.

The variables being multiple, I have not tried to adjust these differences in statistical analysis as suggested by Steiner et al. (2009). These are described in open detail and interpretation is left to readers. Pilot studies are useful in developing the research design and ensuring reliability, therefore a pilot was conducted. Some information was obtained on the measures and the concepts of participants regarding the potential importance of the change in syllabus that the workshops were ostensibly targeting.

Details of the Sample and Missing Data are included in Appendix A2b.

3.24 Instrument Effects

The pre-post-test instruments contain no reference to, or content, that could inform a perception of agency, the independent variable. Instruments measured the change in concept only, the dependent variable. The pre-post instruments could have an effect on enhancing awareness of epistemic issues. Therefore, these were submitted to all in both groups, at the outset of the workshop and as the last task, in the same class room circumstances. Thus, the possible focusing, interest building, enhancing effect of the instruments was available to all members of both groups at the same time in the beginning of the workshops.

3.25 Focus Group as a ‘A Feedback Session’

At the stage of planning the research, it was discussed by my supervisor and myself that an effort to obtain participants’ thoughts on the process would be useful. In the conduct of the experiment, it became difficult to hold the participants back after they had completed the questionnaires of the post-test on the third day of the workshops to arrange a discussion or focus group. Busy teachers have a practice of rushing out after training sessions are over. As the experiment was being conducted multiple times in three cities and I had limited help, I could not allocate this task to others to do at a later time.

An opportunity arose after the last experiment was over to return to school after about a week and hold what is commonly referred to in teacher parlance as a ‘feedback session’. I took permission from the Head of school to visit and talk to the teachers. As it seemed that they were curious and interested, they agreed and a time was set. This could be called a focus group but it had characteristics that set it apart.

Briefly, the teachers who had attended the Experimental Group sessions collected in an empty classroom. I arranged a circle time seating arrangement and had audio recording facilities at hand. I asked the teachers if they would be comfortable with the mikes and did not mind the recording and happily they agreed. The time was immediately after school but nearing afternoon as school had let off early. There was some hustle and bustle outside as some children and staff were still busy. It was a summer afternoon and warm. The fan was working but made a noise which interfered with the recording. I tried to put it off at one time but it became too warm. The teachers were busy when I arrived but agreed to attend. They were in the process of marking scripts as examinations were in progress in this senior school. We spent about an hour together.

I had copies of their pre-test and post-test questionnaires, which I handed out for them to study for a while to help them recall the session as well as to consider the change in their responses. After they had studied the responses and seemed to have done, I asked them if we could talk about it in turns. They took the

mike from each other and gave a few points each on what difference they could perceive in the responses and tried giving reasons. After this the discussion became general. I kept to the rules of circle time, keeping my own input to a minimum but edging them towards my main interest, which was, ‘what was it in the classroom environment, if anything, which could have helped them to change?’ I kept the question tentative and exploratory. The audio record, though somewhat fuzzy, was of a reasonable quality. The purpose of the Focus Group was to inquire how participants recognized change and to what did they attribute this. A few excerpts from the transcript are placed in the text.

Audio records were planned and made of complete sessions but unfortunately, the recordings tend to range between ‘good’ for the facilitators voice, but ‘often inaudible’ of participants’ speech. The recording of 3 sets of Control and Experimental Groups were made but are not a useful record, being fragmented and incomplete. There is, however, enough material in the records and field notes to have assisted me in this detailed write up.

3.26 Data Collection and Analysis

Data in both numbers and words was collected in stages of the research cycle. This is data regarding epistemic beliefs at pre-test and post-test. Nominal data on characteristics of the population was collected with some reliability, however, it is not used as variables in the quantitative analysis. Significant patterns in the data could have been used to analyse and interpret test results, however the variables are found to be far too many and many are unknown or unreliable such as the nature of prior study. Illustration of a perception of agency and change in concepts is sought in qualitative data obtained through recording of the introduction and parts of sessions where possible. The Focus Group, as feedback session, is a most useful source for description. Participants’ oral evaluation of own learning and the process of learning with reference to perceptions of agency in group tasks and discourse was expected to enhance the descriptions of the process and change.

3.27 Reliability

This is practitioner research and issues of reliability can be argued due to the researcher conducting research on her own students as subjects, Cohen et al. (2007). This was, however, unavoidable as epistemic beliefs about knowledge in history is a new area of study for Pakistan and tutors are simply not available. I planned to off-set this by developing a team of tutors of education, a university teacher of history as expert to co-teach the session with the researcher; however this did not work out. I could not find an expert in historiography available to teach neither could I afford to take such a person on the extended round of cities to be present for all sessions. One teacher educator *as* observer was invited, who did attend the sessions in the first round but she did not continue or give me a record of her observations. Another observer was most appreciative but not very well informed. I had therefore, to decide to conduct the training myself. In order to address issues of reliability, a framework, the CBKH, was developed to make assessment of 166 response sheets as reliable as possible. Inter Rater and Internal Evaluation was conducted.

3.28 Inter Rater Reliability and Internal Evaluation

Inter-rater reliability was tried for the BM data assessment. This was followed by internal evaluation conducted in three iterations improving the scale till an acceptable reliability of assessment was achieved (Appendix-A7). Reliability is also tested through statistical analysis and described in the Chapter on Analysis. This is described in Section 3.37.

3.29 Ethical Issues

I, the researcher, was at the beginning of the research, a member of teacher education staff of the organization where the sample was to be obtained but I had no administrative role other than the course I taught. (Teachers currently enrolled

were not to be invited). Before the interventional workshops, I left the organization; however, they were kind enough to allow me to proceed with the research. I also expanded the sample to include teachers from other school systems teaching to O' Level. These organizations have an interest in staff development but the focus of this research, abstract concepts and perceptions, play no part in appraisal of staff at present or in the foreseeable future. The certificate awarded was one of participation and did not indicate success or failure or other information regarding an assessment of beliefs.

Obtaining personal information about age, prior study, nature of study, schooling, was important but I found that teachers hesitated to provide such information and this created some tension so I dropped the idea.

A video I had made on a school visit to an early years classroom is included. This is not used for data but as an illustration of situations we see in classrooms where children are involved in group tasks. Permission has been taken from the School and teacher. As this video was made at least 6 years ago, the children have grown up and moved out of school therefore, it is not possible to take permission from them.

Providing complete information to participants about the research focus (perceptions of agency and epistemic beliefs) could have affected reliability. Reasonably informed consent was therefore, obtained at all levels including each and every individual participant. Informed omission rather than deception Kimmel (1988, p.75) was the strategy. After the research, explanation and individual results will be made available to participants in confidence.

3.30 Some Steps That Were Taken Are

- An intermediary teacher to explain and ask for participation. In the culture of schools, outsiders are not allowed in or allowed to roam freely, especially in current security concerns in Pakistan. Someone who has access to teachers would have to fit the acceptability criteria anyway and there would be 'gatekeepers' controlling that. A teacher from another school branch

or a teacher of another subject area or year group was requested. Opportunity was provided during informal tea time or lunch sessions to answer queries as far as possible.

- Complete security was provided when data on characteristics was gathered. Sealed envelopes, collection boxes were placed for the submission of completed questionnaires. In place of names, code numbers were written and the codes secured on my personal computer. An added measure was to codify names of cities where the research was conducted as well as the names of school systems.
- The pilot had revealed issues in collection where some respondents handed over filled questionnaires to heads in outstation schools. In the actual research, the questionnaires were collected by myself on site to preserve security.
- Guidelines to assist respondents in understanding confidentiality were included in the introduction.
- Teachers need to be well aware of their right not to participate. Cost and benefit need to be explained in the covering note as well as by the intermediary. This practice may not be reliable at all times and contexts, as teachers may feel constrained by their school administration to attend training.
- Efforts were made to negotiate time and venue to be comfortable and non-threatening.
- Participants were free to leave the research at any time. It was explained that no personal or professional consequences for declining an invitation to participate were anticipated. The research was a private project, not funded and no reporting was to be carried out to the organisation. To obtain a formal Certificate of Participation on the workshop, however required participation (Appendix A2a).
- Strict procedures were to be followed in storage of data at all times as well as in the process of analysis.

- Permission to audio-record was taken each time after explanation of the process.
- All records were kept under security. Writing references to the record was done carefully to maintain confidentiality.

3.31 Development of Knowledge and Skills

An interesting phenomenon, change in my own epistemic cognition over the time it has taken for the research plan to evolve and the processes that have fostered the change, is worthy of mention. I have reflected upon this in various sections of the dissertation.

3.32 Description of the Quasi Experiments

Analysis of the Elements in the Design and Threats to Internal/Construct and External Validity in the Contexts of My Research.

3.32.1 Introduction

Cook and Campbell (1979, p.6-8) describe quasi experiments as experiments that have treatments, outcome measures and experimental units but do not use random assignment. Thus, Experimental and Control Groups may be *non-equivalent* where field conditions constrain possibilities of randomisation. Cook and Campbell (1979, p.15-34) use the term validity to refer to 'the approximate truth of an inference' and lay out procedures that they suggest can support assumptions of internal, construct, and external validity in making claims regarding causal relationships between variables. Cook and Campbell (1979), Shadish et al. (2002, p.34) recommend modesty and use of the terms 'tentative' and 'approximate' in making truth claims as 'one can never know what is true'. Steiner et al. (2009) assert that no current quasi-experimental method...provides as convincing a causal counterfactual as a randomised experiment. Where this

cannot be conducted, they recommend a non-equivalent control design with close matching or sophisticated pattern matching. I used the theory to design, conduct the experiments and analyse data, therefore, it is important to take each principle and describe the steps taken in the light of constraints and opportunities.

3.32.2 Truth Claims in Quasi Experimental Research

Cook and Campbell (1979), as self-described 'critical realists', tread the ground with care as they develop their logic of experimentation. They are careful to state that 'outside variables will always impinge on a dependent variable making results sensitive to forces outside the theoretical system' and that they expect the relationships to be fallible and probable at best.

The design of my inquiry is quasi experimental and mixed methods are employed in various ways to collect and analyse data. In this Section of the Chapter on Research Methodology, I discuss particular features of the quasi experiment as conducted in my study in the light of theoretical arguments. Practical problems that arise in research in the field were thus considered in advance and ways to adapt to, and resolve these, were sought. In this Section, I systematically raise threats that can arise in the experimental studies and consider the limitations to interpretation that are suggested in the literature. A balance between reliability and validity is sought through careful design as far as the contexts allow. Rich descriptions are provided with reference to threats to validity and reliability to help readers to interpret the findings of the research as it was conducted. For myself, I draw conclusions on the basis not of despair but pragmatic understanding that field research is not laboratory science but it is the best there is in the circumstances.

Shadish et al. (2002, p.24) quote Cronbach that most challenges to the causal generalisation of an experiment typically emerge *after* a study is done. That is probably true; however, using the copious material in Cook and Campbell (1979), Shadish et al. (2002) many threats to various types of validity were avoided by design in my study. Steiner et al. (2009) suggest that it is better to anticipate a threat before the study where that can be done but if design controls cannot be put

in place, the best alternative is to measure the threat directly to see if it actually operates and if so, conduct statistical analysis to examine if it can plausibly account for the obtained cause and effect relationship. They endorse the direct assessment of possible threats, whether done using qualitative or quantitative observations.

In the studies conducted for this research, the variables, despite planning, turned out to be numerous. By careful reasoning, I was able to plan to counter some threats. Others that could not be countered are described as well. Assessment of each variable would require reliable evidence, and adjustment to the numerous variables is difficult, and there are areas where evidence is not available. My effort is to describe conditions with transparent detail in order for interpretation to be well informed.

Shadish et al. (2002, p.40) recommend the primary method of ruling out threats is to use design controls that minimize the number and plausibility of these threats that remain by the end of the study. As Shadish et al. (2002, p.484) term it, fuzzy probability is central to ruling out threats in quasi experimentation.

Cook and Campbell (1979) discuss two classes of designs; those they consider generally non-interpretable, that generally do not permit causal inference, and non-equivalent, control group designs that, they say, are generally interpretable. Among the latter, the untreated control group with pre-test and post-test is most often used and often interpretable. Cook and Campbell (1979, p.103), Shadish et al. (2002, p.136) This is the design I have used in my studies. Groups were not matched in particular. Broadly, all belonged to the category of O' Level teachers of teaching history in high school. They are therefore, classed as nonequivalent.

What follows in this Chapter is a description of ways in which threats to internal validity were countered during my research. This is structured like a list with separate headings for easy reference. The theory is raised in the preceding Section.

3.32.3 Ambiguous Temporal Precedence

Shadish et al. (2002, p.55) reiterate that cause must precede effect but consider this may be unclear. In my experimental research the causal treatment was provided between pre-test and post-test in all 5 studies. Measures were carefully conducted at outset and in the last hour of the final day. There were no other training sessions being held to develop epistemic beliefs about knowledge in history in schools or at CIE or the British Council in Pakistan at the time. The chances of this being a valid threat for the sample are few.

3.32.4 Selection and Assignment for Control

Research participants for this study were found in three large private school systems in Pakistan spread in three large cities Islamabad, Karachi and Lahore. These school systems prepare students for Cambridge O' and A' Level examinations. The target population was the entire population of teachers of O' Level History in all these school systems. An estimate given by System One, which is the largest, of the number of history teachers was 200. However, at the end, only 83 were accounted for after data cleaning. The other schools did not give me an estimate and only offered vague responses over the telephone.

Potentially, the complete population of adult O' Level history teachers between Classes 9-11 in all 3 school systems was invited and therefore, could have attended. In that sense I can claim that there was no bias as there was no limitation and no selection on my part. No one was disqualified due to any reason by me, the researcher, or anyone on my team. Most did not attend however, upsetting my original calculations. This attrition took place at the outset of the programme and there was some attrition during the programme. This is documented and will be accounted for in the analysis. Reasons for not attending are probably random. Discussions in the pilot prior to conducting the research had disclosed that teachers did not consider that the prospective change that I was

targeting in the curriculum was of much consequence. Cambridge had indicated this in writing on the website.

The largeness of the school systems and the lack of access made it difficult to judge if self-selection, administrator's assignment or other reasons were behind the reduced attendance. Crucially, as self-selection itself could be considered as contributing to the outcome of interest, bias would be an issue. A perception of agency would be compromised if someone had been ordered to attend. There is, however, one argument which is important. As these are Senior School O' Level teachers, the school culture is not authoritarian for them. O' Level teachers are greatly in demand in the city and often work part time moving from school to school. School administration would be more likely to persuade than try to force attendance in a three day workshop. There is very little professional development offered to senior school teachers. Even CIE has about one session a year, which is expensive, therefore, schools select some teachers to attend. There would be some value seen by teachers and schools to develop their practice and add training to their CV.

Groups were in one sense given, as school managements made the arrangements to their own convenience. These were not intact groups such as a complete university class but yes, in the sense that these were, as history teachers, a subset of all the teachers in the school system. I had distributed posters to schools to encourage volunteers but knowing schools, I can conceive of a range of ways in which teachers were either told to attend or heard about it and requested to go. A scenario of random selection of teachers in busy functioning schools is hard to imagine.

As in some cases, out station teachers attended, there were variables that were not known such as travel issues, weather, time pressures, etcetera. These are necessarily the kind of differences that would underlie the makeup of groups and social situations. Homogeneity in matched groups would still have hidden differences. A sick child at home, transport problems, copies to correct piling up, could not possibly be matched or accounted for.

3.32.5 Assignment to Control and Experimental Groups

Assignment to Control and Experimental Groups was also due to circumstance and not design. Teachers' presence in the geographic area, a matter of chance, made it convenient for them to attend either one of the two groups formed in each study. To explain this further, either their home or their workplace was in the location of the workshop. In three cases, there was switching from control to treatment group on the second day. These respondents are not included in the analysis.

One other way of control was that I determined in advance that all Group A's (the first in the set of Group A and Group B) would be Control Groups, therefore removing any possibility of choice on my part as to which group should receive the treatment. The groups were constructed by the school to my broad specifications. I could not influence who would attend or in which group. That this meant that the experience would be different between groups in any way was not known to anyone except myself and my supervisors and the ethics committee as earlier stated. Being part of an experimental research is not commonly experienced here nor do teachers know that an experiment would mean different experiences within groups. Therefore, it is safe to assume that only practical, personal reasons were behind the assignment of teachers to either group.

3.32.6 Differential Attrition

Differential attrition is hard to judge as members of groups attended, or did not attend, due to random, unknown, administrative or personal reasons. Steiner et al. (2009), in the context of experiments, suggest that if the pattern of attrition differs by group then a selection confound is introduced. There is some clear pattern for differential attrition that can be identified in one of my studies. In the first Control Group, 16 participants were reduced to 10 on the second day, and went up to 11 on the third day i.e. 5 did not submit a post-test. In the Experimental Group of the study, there were only 6 on the last day from a total of 12 on the first day.

Thus both groups showed attrition. In this city study there was a concern expressed by participants that the change in the syllabus was of not more than 4 marks, therefore learning about assessment of source did not seem of much value. This could be a reason. Another Control Group of Jasmine lost 4 participants at post-test while the Experimental lost only 1. In this group there were several travelling from out station. An important point is that the first three studies had participants of one school system. The other studies had participants of two systems. The remaining studies did not show attrition in the same way with only one or two not attending either pre-test or post-test in the Control or Experimental Groups. Pansy, the last study, has attrition of only 1 participant; however, change is seen as more in the Control Group rather than the Experimental Group in this study. Such participants were counted out of the analysis. Travel from distant towns should not be discounted as a reason for attrition. Missing data does not include these participants.

Regarding the aspect of threats to the interpretation due to differences in voluntary or involuntary participation, I can make some clarification here. I, as researcher, had a very limited access to schools and was dependent on school administration to contact teachers and make arrangements. While schools were sent the poster and told in writing that the workshops needed to be voluntary, it was difficult to know how voluntary this was in real terms. It can be suspected that schools' arrangements differed in, informing/inviting/persuading/even charging teachers to attend the offered workshops. As there is a dearth of professional courses for teachers of senior schools, it is possible that the workshops were welcomed. A record of the last feedback session provides useful data in this area. It might be better if this study is repeated to obtain data about motivation and intention at the outset of the course.

3.32.7 Access and Social Situation

I did not have free access to schools and was only permitted to visit when conducting the workshop. This limited possibilities of randomisation, selection,

assignment, location, and even the kind of observation that could be conducted. Most researchers would face such problems of limited access to sites in my opinion.

As Shadish et al. (2002) describe, each experiment is also a social situation, full of social roles and social expectations but with a uniqueness that can lead to problems when social cues are misread or thwarted by either party Shadish et al. (2002, p.30). I experienced this in the very first group which happened to be a Control Group. I found a trio of members that were considered on the top of the field in history teaching in the flagship branch of the School. Good results of the School and long term experience made them feel any further training to be a waste of time. They had come along as I was a known trainer in the field and they probably did not want to miss anything but they were looking for specific examination related content. The focus on the change in history papers did not appear useful to them as they were not convinced that this would mean a larger change in examination content. I was told by one participant that it meant a mere 4 marks difference in the examinations. (Early data gathering had established this perception existed among experienced O' Level teachers of history. This was the mark difference that could be worked out from CIE guidelines)

One participant, especially, was a part time teacher in the School and a visiting teacher in other schools with little time to spare for flighty ideas about history content. She was a constant source of trouble in classroom proceedings, challenging me on each mark awarded when we did some paper moderation. Such conditions, neither desirable nor replicable, are albeit, part of the complexity of field settings. It is difficult to quantify such an experience to add to statistical analysis because of its unique and complex nature.

Another example that is important is of a young teacher who drove down from a neighbouring town to attend classes in the city in spite of bad roads, insecurity, and a distance of over 120 kilometers. Most teachers that come long distance do so in school transport and do not drive alone. She persisted in coming even when she had an accident. Such examples within a group question the level

of homogeneity even with all other things being the same. Interestingly, she was in the Control Group and grew in degrees of sophistication as assessment shows.

3.33 Alternative Explanations

3.33.1 Prior Study

In this experiment, one alternative explanation for an initial score in a pre-test could have been background of study of history as a discipline in a degree programme as opposed to a degree in other subject areas such as Pakistan Studies, where processes of history e.g. source assessment are not taught. Information regarding earlier qualifications was asked and, when it became apparent that the domain of the degree was not forthcoming on the questionnaire, I requested teachers at the start of the session to please add it. This was unfortunately a source of some tension as it affected their credibility in the eyes of their colleagues and I could see that they considered it confidential. I therefore, gave up the attempt. This is one of the limitations that make causal inference difficult from weaker quasi experiments.

A difference between study in history and the subject Pakistan Studies has implications for epistemic beliefs but obtaining this information reliably was difficult and seen to be embarrassing, therefore I discontinued it. I felt that I may be crossing limits of ethics in trying to obtain data they were loath to give. I should in principle work out the effect of such variance and account for it in analysis. This variable could make a difference but information regarding the variable is not available and would be a contravention of ethical concerns to extract.

3.33.2 Knowledge of the Treatment

There are no guarantors of valid inference argue Shadish et al. (2002, p.483). In my research, knowledge of the treatment would have had an effect, therefore deception was used to control the threat. Deception is difficult in practice

as well as to justify especially in a research where agency is being critically explored. It was necessary, however, as explained in the Section on ethics, due to the importance of studying the relationship and the unusual circumstances that made this research design possible.

All groups, Experimental and Control, received one level of the treatment, which was instruction to change concepts of knowledge in history. The method of instruction was Intervention 1, which was an experience of a range of suitable examples. No one was aware of the existence of or the nature of the underlying concurrent treatment to create a perception of agency except for myself the researcher/teacher and my supervisors. This treatment was administered to Experimental Groups alone as was expected to effect the change of concepts making new ideas more acceptable.

3.33.3 History and Maturation

An issue in such experimental studies is the threat of change emerging from learning conducted elsewhere. In-service learning tends to filter down in schools with small workshops and collaboration. It was, therefore, necessary to frame the research to short courses and complete the experiments before some other training event took place. As this was not an award bearing course, assignments, and long duration study would not be acceptable to a large sample anyway.

This was a one in a million opportunity to study a planned and necessary change in concepts within a bounded population and I did not want to let it go to waste. The threat that history could possibly explain findings of change, both possible and plausible, drove the pace of my research programme as well as choice of instruments. The History O' Level Curriculum of Cambridge International Examinations (CIE) was in the process of change and it was appropriate to expect schools in Pakistan and CIE to institute teacher education programmes to address it. I feared that CIE could be conducting workshops, or even, that teaching practice to the newly developing curriculum itself could affect teachers' concepts given time, thus becoming a threat. I chose therefore, to use existing, validated,

instruments rather than wait to develop and test new ones for Pakistan's context. I planned the workshops to take place within a short space of a few months in order to pre-empt other training programmes making a difference. These were both possible and plausible threats that I was able to successfully counter.

An alternative explanation for change in concepts could be earlier or simultaneous attendance by the participants at a teacher training programme with similar content. As such programmes on this change in curriculum or related subjects were not being offered in schools, CIE training could be the only other possibility (I did enquire). Before beginning the workshops, I had attended one such training by CIE to find that the training sessions did not target teachers' concepts about knowledge in history. I held the workshops before training was conducted by CIE again that year, and gathered information regarding those who had attended the training earlier. I held meetings with a group of teachers who had attended training sessions and was able to conclude that the concept of source in history was assumed to be only a 'hint or link' to assist in answering questions in examination. A key way to reduce the threat was to take pre-test and post-tests and time sessions in such a way that no other training session was held or attended between these. This I was able to ensure, in the sense that I kept a track of any such training on offer within the time scale. Luckily these did not take place.

3.33.4 Threat of Testing

The population of O' Level teachers of history being specific and small, made the threat of testing real and thus limited the extent of the pilot that could be conducted. Those who had studied the test in advance could not be assumed to respond in comparable ways as those who had not. I tested the instrument EBI on random family and friends, colleagues who were not from the target population, and the history instruments on some teachers who were then not included in the experiment. As numbers mattered, this was a difficult decision.

As far as the pre-test was concerned, it was not a threat because it was administered to both groups, Control and Experimental, at the start of the workshops and at the end.

3.33.5 Resentful Demoralization

An issue with withholding treatment did not occur as both groups received a matching content in instruction targeting the stated goal of curriculum revision. The difference of treatment between Control and Experimental Groups was qualitative as an enhanced perception of agency. Thus practical issues of groups learning different content were avoided and there was also reduced risk of resentful demoralisation.

One issue which Cook and Campbell (1979) refer to is 'Resentful demoralization of respondents receiving less desirable treatments'. Steiner et al. (2009) similarly refer to the Stable-Unit-Treatment-Value-Assumption (SUTVA) where it is advised that the experiment is conducted in such a way that students and teachers do not become aware of the specific treatment or control condition to which they were not assigned. The treatment was at two levels. Both Control and Experimental Groups received Level 1, which was the course itself, to address changes in the O' Level History Curriculum. The experimental treatment, the perception of agency, was given only to the Experimental Groups.

All participants knew that I was carrying out research and had signed consent forms, but most had never taken part in an experimental research before and did not know that different treatments were to be given or expected. The statements I made regarding agency, my efforts to encourage disagreement and challenge, may have appeared my mood of the day. This threat was therefore, assumed not to apply within the short time frame of three weeks within which each study took place.

3.33.6 Threat of Difference in Experience with Different Tutors

Another threat, of difference in experience with different tutors, did not apply as I conducted all sessions myself. There were many teacher educators available who teach education. Tutors were required with matching knowledge, skills, and understanding and critically, perspectives at a philosophical level, as was needed for this course to be taught. Such tutors were not available. Course objectives were to change teachers' concepts about knowledge in history. A tutor for this course needed to possess a working, practical sense of epistemological issues, and of status of sources in knowledge in history. As this was described as practitioner research in the original brief, I could decide to keep the role to myself. Though extremely taxing on my time and energy, this threat did not become, as it could have, an intervening variable across all six studies.

The researcher being the one delivering the treatment was, however, a threat to inference as being the researcher and knowing the Experimental group. I could have unconsciously or consciously made a difference. What difference could that have been, to obtain a change in epistemic beliefs take place amongst more participants, in the appropriate direction and degree, and match it in all 5 studies might then well be the factor of interest! There is a clear difference of more change in 4 out of 5 Experimental groups and a perception of agency is the difference I consciously maintained. If there was any other common factor, I do not know of it.

3.33.7 Size of Sample

Initial questioning in schools indicated that the sample could be large enough for inference but unfortunately, a smaller than planned sample was obtained. The reasons could be many. Concerns of safety and security in the country affected all equally and made a large difference in reducing the numbers that could attend. Those for example in the north could not attend.

3.33.8 When Treatment is Non-Obtrusive or Comparison Becomes Difficult

Cook and Campbell (1979, p.348-349) discuss obtrusive and unobtrusive treatments where respondents may possibly not notice the treatment at all or pay scant attention to it reducing the effect of the treatment. This could certainly be an alternative cause in my research studies. Cook and Campbell consider such treatments could be mundane, or not stand out. In the case of this treatment, I can tell from experience of conducting this experiment that the treatment does catch attention. It is palpable in the classroom atmosphere and the qualitative data does provide evidence of this. Cook and Campbell (1979, p.348-349) do suggest multiple research programs to counter this threat and that is good advice. I carried out the experiment 5 times, which helped to make my observation, as stated above, somewhat more dependable.

3.33.9 Non Availability of Baseline Assessment

Cook and Campbell (1979, p.348-349) point out problems where a no-cause baseline is not available for comparison as some level of treatment was given to all. Cook and Campbell refer to this making it difficult to test hypotheses about 'absolute cause' as opposed to hypotheses about differential impact. Campbell and Cook suggest a no-treatment Control group would solve the problem. In this study, that would not help, as to create a workshop situation and not target the curriculum revision, would not make sense to schools or teachers and neither would any other training correspond to the domain of the measures. Tests to obtain no-change data from those who did not attend any training would still not help in differentiating between Control and Experimental Groups. The best option in this case is the one employed, that of Control Groups as well as pre-test and post-test measures. In order to add to reliability, I designed each study with a Control and Experimental Group. Another important factor was taking entry pre-tests for all groups. This has been a most useful element as a comparison could be made with some control over variables. Both groups had participants from the

same geographical area and the same school system. The pre and post- test for each study is a useful device for comparing change. Out of 5 studies, 4 showed more change in the Experimental Group. One group differed with more change in the Control Group.

3.33.10 Confounding Factors

My greatest concern was, if the design of the intervention was too strong, as it appeared to be in the pilot, it would explain the change making the perception of agency of little effect. Both groups would perform in similar ways. To explain, the content of the workshop, History of the Sub-Continent, as part of the teachers' taught curriculum in schools is an emotive subject on the whole for the general population. A choice of the massacre at Jallianwala Bagh used in the pilot surfaced personal feelings that underpin the colonial experience. I feared that this would make agency of little consequence. As it turns out, there is good change in both groups, Control and Experimental but more change is evident in the Experimental Groups.

I structured the experiment to ensure that both groups experienced the same content, topics, and questions. The perception of agency was a treatment reserved for Experimental Groups alone. I figured that if the difference was seen over all iterations of the experiment, it could be plausible that agency had indeed had effect over and above the effect of the content which was common to both.

3.33.11 Construct Validity

3.33.11.1 Inadequate Pre-Operational Explication of Constructs

Concepts, conceptual change, epistemological beliefs, beliefs about concepts in history, beliefs about history, are discussed in depth in the literature and reviewed in this study. Cook and Campbell (1979, p.65) suggest that a precise explication of constructs is vital for high construct validity as it permits tailoring the manipulations and measures to whichever definitions emerge from the explication.

Concepts and conceptual change are defined in various ways by scholars. Precision or even consensus in definition is not to be found in the literature. Formal definitions are therefore, those that pertain to the measures. I was fortunately able to contact the developers of two of the measures to obtain permission as well as understand ways in which these were used. The measures being common to both groups did not have differential impact.

The manipulation, as mentioned, was at two levels. Instead of no explicit treatment to the Control Group, the treatment given was 'qualitatively different' (term used by Steiner et al. (2009). At one level, all groups attended workshops in education to change concepts of history. The content and process was designed carefully following VanSledright (2002), VanSledright (2011). Three different examples from history, as written in relevant course texts, were selected and materials were organized to help participating teachers experience examples of how history is represented. A range of sources was examined to provide experience of evaluating source. This was a careful match and arguably met standards of validity.

Shadish et al. (2002) argue that since single operations both under represent constructs and contain irrelevancies, construct validity will be lower in single exemplar research where each construct is operationalised in multiple ways in order to triangulate on the referent. They suggest that there is no substitute for deliberately varying two or three exemplars of a treatment, where possible. This I could not do.

Definition of the construct of agency was not as large an issue as agency itself was not being measured. The perception of agency could be described with the use of qualitative data from the sessions. The audio record of the sessions helped to flesh out my own notes to build an account.

3.33.11.2 Explication of Treatment

Careful explication of the treatment can be possible. The independent variable, a *perception* of agency, was explicated in the Experimental Group and

this same explication used each time. Thus the treatment could be planned and maintained over iterations. How this perception was perceived, could also be examined with the use of qualitative data from the sessions. The audio record of the sessions, my field notes, photographs, helped to build a description. It is my intention to suggest ways this can be researched in the future with different exemplars of the treatment. There are, however, differences in degree of explication or perception that must reasonably have taken place and these are not measured or indeed even known.

As Shadish et al. (2002) note that construct validity is fostered by, one, starting with a clear explication of the person, setting, treatment, and outcome construct of interest, and two, carefully selecting instances that match those constructs. While efforts were made by me to maintain the pattern, curriculum, school settings, sample similarities between iterations, there were differences. These are enumerated. The balance between educational qualifications, age, sex, experience in the units studied was different in each study making exact replication impossible to claim. As I have mentioned in the Section on measures, I did try to look for a ready instrument to measure a perception of agency; however, the construct being new, ready measures were not available. There were none that precisely targeted a perception of agency in task based group discourse. There was not enough time to develop and validate new ones as explained. A measure such as this could have helped to address this threat with more reliability.

3.33.11.3 Variables in the Sample and Settings and Tutor

The letter promoting the course described the target audience thus the target population of teachers of O' Level history was defined and met. In one school, several teachers of a lower class and age group (8 rather than 9 onwards) and teaching a different text book, attended, who had sadly to be counted out in the analysis. Some attended in a group who were teaching Matriculation rather than O' Level history. They were also not counted in the final total of 83 respondents. The measures remained the same in all studies but it could be

argued that participants would respond in individual ways to the language, content, perspective, task demand. Settings were schools or their teacher education centers. These are described. All settings were similar in the sense that attending teachers were familiar with both location and routine of teacher education.

The settings were as similar as field experiments can have. There were differences that are described. One was noisier, another had an idyllic setting but was in the school group office training room. All sessions were conducted in school training rooms, I was the tutor for all, one group was smaller as members did not turn up. The training was conducted in three cities, thus there were cultural and language differences. However, these differences were between iterations or studies. In each case, Control and Experimental Groups had similar settings and broad person characteristics. In each case, the group was from one homogenous background of school with 'surface similarities', Shadish et al. (2002, p.67).

The quasi experiment was constructed in a sensitive and careful manner, recognizing constraints and opportunities of the context. Participants were not treated like laboratory animals and an ethos of shared endeavour, everyday work place activities and camaraderie was maintained. The process was organized to create Control and Experimental Groups, pre-test and post-tests were taken, the treatment was limited to the Experimental Group and the whole process was repeated over six iterations of studies in three different cities of Pakistan. Both qualitative and quantitative measures were deliberately selected to triangulate and expand findings. As the measures were taken from studies conducted outside Pakistan, and the questions covered an unfamiliar domain, it was seen as necessary to use both closed and open ended structured and semi structured questions.

Was the measure a perfectly valid reflection of the theory being tested? It was hard to say as the instruments were piloted on a small scale. The data from the pilot was inadequate for statistical inference. For one, a large scale pilot was not practical because of the smallness of the population of history teachers, making history a threat if the instruments were experienced repeatedly. Time constraints were an issue as earlier described in the Chapter on Research

Methodology. A careful review of the responses demonstrated no problems and so it was decided to go ahead. A post-doctoral professor in Psychology, who is an authority on research in Pakistan, was asked for her opinion and she found it suitable for use. Small notes were added to explain the meaning of a few terms on the questionnaire itself.

In my study, I carried out pre-tests in both Control and Experimental Groups as I was sure that a prior experience of the questionnaire would give an advantage to the respondents and become a possible alternative explanation. Besides this, in each study, I maintained both Control and Experimental Groups because there would be differences such as language skill, familiarity with training, school ethos, etcetera, that could account for the effect.

The teachers participating in the studies had not received any treatment to change their concepts about knowledge in history previously. The treatment was thus at two levels. On the one level, observations could be made about being better prepared to teach the developments in history curricula. A form of knowledge and skill addition was the intervention at this level. The manipulation was the steps taken to enhance a perception of agency in the Experimental Group alone. This was in terms of words spoken and steps taken to encourage disagreement, challenge to the teacher’s perspective as well as that of the group. Encouragement to explore one's own prior understanding within group talk was less well understood as teachers were not used to the idea and would require a different kind of training. This is further explained in the analysis.

In my experiment, as explained earlier, I was the tutor for both Experimental and Control Groups in all cases. I was the one who implemented the treatment and withheld it in Control Groups. This had advantages in the sense that I knew the treatment and the centrality of the approach in classroom practice. I could carry out the steps designed by me and follow the procedure as planned as well as vary my approach. There were disadvantages. The obvious issue of bias is one that can only be dealt with statistically, if at all. This was unavoidable as described earlier; however, the threat of bias is real and has to be accounted for.

3.34 Variables in the Treatment

3.34.1 Slip Between Intention and Action in Classroom Practice

A slip between intention and action in classroom practice is possible as years of teacher education and classroom observation as well as much theory has taught me the difference between teacher intention and action, and teacher intention and student perception. How often teacher instructions fail to be heard, understood and followed, is known by all those who spend time in teacher observation. Perception games are part of teacher education's bag of tricks. I do not assume that the group of 20 or so adult teacher participants perceived the same message equally and it made the same impact on them in every one of the six groups. However, having said that, research in the social sciences has to be conducted in the conditions that exist. That is also the only way research can have validity. If I were to tell my teachers I had conducted this experiment rigorously, in a laboratory, and these were my findings, they would say alright now do it here in a class of 15 year old boys and a curriculum for examination.

Unlike medical research, where implementation requires unobserved actions by recipients, who are out of the researcher’s control, this was in a sense, easier because I was myself the person implementing the treatment. In a follow up research, I would recommend that student logs be instituted to observe if the treatment has been similarly perceived. I would also suggest a video of a presentation be added for Experimental Groups to ensure that main features are communicated and this can be better defended. I would also suggest that there should be checkup points for the person implementing the treatment to self-monitor the process as it unfolds.

My experience of years of teaching and teacher education, and my own observations during the programme of research tells me that they did perceive the agency. What they thought was this was my style. The way I challenged them to think on their own and contradict or differ with me or their group, *did on the whole*,

make an impact. The feedback session is good data on this. Snippets of the audio record of the sessions also points in this direction.

I tried to institute observation for the sessions but the observer I had asked did not show up on the second day of the workshops. I had other occasional observers from the schools who attended some random sessions but they could not be told what the treatment was as this was a planned deception. Observers normally take notes on school observation formats they are familiar with and a perception of agency not being one of the items on these formats, one observer I had requested, did not record or communicate anything. The random observers did not sit through complete sessions and certainly did not sit in both Control and Experimental Groups.

I had no one but myself to monitor my actions and record anything. The audio records were not continuous as electricity is not continuous in Pakistan. Technical hitches, sound variance affected the record but in spite of this, I can say I did make the required statements, and did persist in the planned approach. How many student teachers heard, recorded, and developed the perception, was not recorded or measured.

Having said this, I can say that there was a perceptible difference in the classroom environment between Control and Experimental Groups and I have data that pointed to this. The feedback session was the most useful source for such a claim. The transcript of the audio record can support this.

Shadish et al. (2002) cite a well-known maxim, co-relation does not prove causation. They explain that this is so because we may not know which variable came first nor whether alternative explanations for the cause exist. They add that co-relations also do little to rule out alternative explanations for a relationship between two variables, Shadish et al. (2002, p.7)

Quasi Experiments are Quasi Scientific conducted in field settings and not laboratories. In educational experiments, the field setting can have a classroom with students, teaching practice and normal testing conditions that has little to distinguish it from normal practice. Indeed, as many concessions are made to human subjects and environments, causal inference is tentative. I did explicate all

possible threats to inference and design ways to deal with them but it was difficult to know if this was an exhaustive list. Collecting data to rule out effects is problematic at best. I have made efforts to counter threats in design as described.

I have also used mixed methods. I thought that a good description supported by evidence could be built with both quantitative and qualitative data complementing or even arguing with each other. Thus the need to use both approaches seemed both natural and necessary.

The classroom experiment, or rather the series of experiments, provided valuable contexts to study the miracle of changing concepts. The tests themselves have descriptive potential and the field data provides rich description to talk about the change and how it took place. At the outset I wondered if, as Meehl (1990) argues, this is one example of an interesting and important idea that is really untestable. I decided that I would reduce it to practical situations and processes in order to test it in my own practice. I have not tried to prove or falsify the theory, just learn enough to build a hypothesis for future research into conceptual change. I may not succeed in this instance but it would be worth doing anyway. By quoting Meehl, I demonstrate my awareness of the problems of testing theories in the social sciences. As a teacher researching my own practice with quasi experimental methods, I hope to claim I have enough evidence to argue that this interesting intervention may have made a difference to the learning and can possibly be researched.

This Section of the Chapter on Research Methodology discusses the quasi experiments in the light of theoretical arguments. Transparency in description is vital as the alternative hypothesis is seen as being supported with the findings in this research. The findings are otherwise reasonably reliable and effect sizes suggest a moderate support to the hypothesis. More detailed knowledge of how the research was conducted in the field will enhance the interpretation. An organizational feature is used in this essay for easy reading. References are made to theoretical arguments and each is discussed according to how this threat was addressed in the limitations of the context. By focusing on an explication of threats and describing how threats were dealt with in my research context is useful to

interpret findings. As a practitioner in the field, I can thus demonstrate how far threats to reliability can be countered in real terms.

3.35 The Measures

This educational research is planned as a series of experiments in the field with the use of mixed methods in data collection and analysis. I needed to select appropriate instruments that could measure the same construct in different but compatible ways to obtain as reliable and valid an assessment as possible. This Section discusses the Measures with some reference to theory as well as practical constraints and opportunities. The qualitative measure is a set of open ended questions called the Boscolo and Mason (2001) questions (Appendix-A3). Two, Likert Scales are used. One is the BLTHQ, Beliefs about Learning and Teaching history questionnaire (Appendix-A4) Maggioni et al. (2009a), and the EBI, Epistemic Beliefs Inventory (Appendix-A5), Schraw et al. (2002). Apart from these, data from a feedback session was audio recorded and field notes were made which are used to describe and discuss the proceedings as well as to illustrate the findings.

3.35.1 A Measure of a Perception of Agency?

In order to decide the measures to use, my first consideration was if a perception of agency, the independent variable, could be measured. It would have been convenient to measure the perception and co-relate the data with a measure of belief change. There are validated instruments available for self-esteem, self-regulation, and locus of control. Psychological tests, e.g. Levenson's Locus of control scale; Sherwood's Self Concept, were studied for this purpose. An instrument called the CAMI[38] measures agency, control and means-ends beliefs of children and adolescents in school settings. These are however, instruments to

[38] CAMI Control Agency and Means-ends Interview (Skinner, Chapman, and Baltes (1988))

study individual qualities rather than the effect of environment or social context. My research does not focus on stable traits, beliefs or attitudes.

I saw that an appropriate instrument to measure a perception of agency in a social context was not available readymade and developing a new one would have been time consuming. As time was a key factor in the change process, the research was designed to measure the change of concepts and not a perception of agency.

For the 'perception of agency' as an intervention, I assume that the process of information-communication regarding individual agency has been effective to enhance the perception amongst participants in the Experimental Group. In the light of this deliberation, I considered that a pre-test, post-test, non-equivalent group design Cohen et al. (2007, p.283) was most suitable. Change in concepts would be measured and the perception of agency would remain the treatment and would not be measured. Qualitative data would be collected, if possible in a focus group session, seeking respondents' views on the treatment. Field notes and audio records would enhance a qualitative description.

3.35.2 Questions to Consider in Instrument Selection

Recognising that there are numerous instruments on offer to study epistemic beliefs, Duell and Schommer-Aikins (2001, p.421) suggest 4 questions to consider in selecting an instrument. These are: consideration of theory, (e.g. do instruments measure beliefs as multi or unidimensional); relevant epistemological dimensions, (e.g. justification); format, validity, and reliability. These questions were among those used to make a decision on measures to use. The questionnaire for qualitative data collection is called the Boscolo and Mason or BM Questions. The questions are adapted from those used by Boscolo and Mason (2001) Bozo, Morra and Pierimarchi, described by Mason in Mason (2002, p.328). Responses to these questions were assessed using the CBKH rubric.

3.35.2.1 The Open Ended Boscolo and Mason Questions:

1. What is history?
2. How do people who write history know about the past they write about?
3. What problems can historians have when they try to understand what happened in the past?
4. Is it possible to explain what happened in different ways?
5. Why?
6. If there are two different explanations, how is it possible to understand which is better?

3.35.3 Some Considerations

A pilot was conducted with the addition of historical context to the questions. The addition of historical context was found unsuitable and the instrument was returned to its original form. It was understood that the questions had possibilities of yielding rich data, cause respondents to think and provide information that could help to interpret the data from the Likert Scales.

3.35.4 Measure of Beliefs about Knowledge in History

The BLTHQ was planned as a measure of beliefs about knowledge in history. The EBI was suitable for triangulation with this as the BLTHQ also explores beliefs about learning and teaching and has a Likert Scale format. The BLTHQ tests domain specific beliefs about history and has been through a validation study by Maggioni et al. (2009a) with mixed results. Maggioni et al interpreted the results of the confirmatory factor analysis as supportive of the BLTHQ with the theoretical framework and thus an initial indication of its validity (Maggioni et al. (2009a, p.205). They also report however, that a low number of items and relatively low reliability of the scale indicate that it is still a rough measure. A second version of the scale, the Beliefs about History Questionnaire, BAQ, is described in another study by Maggioni et al. (2009b). The BLTHQ was considered suitable.

3.35.5 The Epistemic Beliefs Inventory

Two direct, self-report measures of epistemic beliefs were selected. One of these is the Epistemic Beliefs Inventory, EBI (Schraw et al. (2002, p.261). The EBI measures domain general epistemic beliefs from a multi-dimensional beliefs perspective. These are stated as beliefs about Certain Knowledge, Simple Knowledge, Quick Learning, Omniscient Authority, and Innate Ability by Schraw et al. (2002, p.263). The EBI is a development from the CLEV, and the EQ. The CLEV formed the basis of the development by Schommer-Aikens (1990), of the Epistemological Questionnaire, the EQ. In a comparison study of the EQ and EBI, (Schraw et al. (2002, p.271) conclude that the EBI had better predictive validity when co-related with a test of reading reliability as well as test-retest reliability than the EQ even though criterion validity remained poor (Schraw et al. (2002, p.272). Findings indicated that the EBI yielded 5 hypothesized factors described earlier. Compared to the EQ, the EBI explained substantially more sample variation than the 5 primary factors on the EQ although neither instrument produced factors that were highly reliable.

Hofer (2002) developed an instrument called the 'Epistemological Beliefs Questionnaire', EBQ, finding a 4 factor frame work: Certainty of Knowledge, Source of Knowledge, Justification, and Attainment of Truth. Schraw and Olafson (2008, p.28) find that the EQ, EBI and EBQ explained approximately 45% of the sample variance, a mixed success. Schraw and Olafson (2008, p.29) express concern regarding current instruments. They find that existing self-report instruments have not agreed upon what should be and can be measured, show low predictive validity between epistemological factors and various outcome variables. They say the instruments measure narrow epistemic beliefs rather than holistic world views, the focus is exclusively on epistemological worldviews without considering ontological world views. They propose a new instrument 'The Four Quadrant Scale' (Schraw and Olafson (2008). I piloted this instrument in a sample of teachers and again it was seen to have for teachers, a strong indication of socially desirable positions. Teachers of O' Level Science, who had never

experienced questions of philosophy, were seen to choose the epistemological, ontological relativist positions. This could be due to the questions being similar in tone and content to the school's classroom teaching manual instructions with items such as 'students need to be involved in learning through discussions, projects and presentations' and 'students work together in small groups to complete assignments as a team'. It is a very good instrument but adaptation would enhance the questionnaires predictive potential.

As pen and pencil measures are still developing, generally, 45% validity is cited Schraw and Olafson (2008). The choice of the EBI was made as it is among the most used instruments. There was reasonable evidence of reliability and validity found in the study by Schraw et al. (2002). The Likert Scale, although of a 5 point, did provide continuous data. The instrument was easy to score and administer and did not require certified people to score it. More importantly, as there is knowledge of its use in various studies around the world, this can be useful in assessing the quality of the data that emerges in Pakistan. There are 32 items as compared to 63 items in the EQ making it more manageable and less tiresome to participants unused to doing questionnaires especially ones that ask odd questions about truth. A pilot was conducted; however, the smallness of the sample, 35 respondents, did not yield conclusive data against the EBI in the Factor Analysis. Therefore, it was decided to go ahead.

3.36 The CBKH Category Scale

3.36.1 Introduction to the Proposed Scoring Rubric: Categories of Beliefs about Knowledge in History the CBKH

This Section briefly outlines developments and arguments that are discussed in the Chapter and presents the rubric, the CBKH i.e. Categories of Beliefs about Knowledge in history. The two main versions are produced here to describe the development and are useful alternatives in their own right. One is a weighted scale and the other is without weights. The 7 Categories are listed and

the elements that make up each Category are defined. Excerpts from the responses to the Boscolo and Mason questions are also arranged in a table to serve as exemplars for use of the scale. These exemplars are useful for a consistent assessment and to understand the differences in responses according to the Categories.

The rubric developed in stages of trial and error. Repeated assessments of the data were made with each developed version and reliability of marking was assessed in several attempts. This process informed the modification and development of the Category Scale. Reviewing the literature and discussions with Liliana Maggioni raised the level of the analysis and reflection on the model. The CBKH was finalized after Inter Rater and Internal Evaluations, provided a reasonably consistent, therefore satisfactory assessment of responses to the BM questions using the Scale.[39]

This Section of the Chapter on measures is followed by a Section with a discussion on the literature in the domain of epistemic beliefs to provide a background and rationale for the development of the model. The Section also contains arguments with the current definitions of the categories and rationale to unravel and simplify the levels for use in a less than expert population of history teachers. It is important to note here that the CBKH is *a* model, one of many, and there is not enough known or agreed upon about epistemic concepts to conclude that this is the most appropriate order or definition of categories. Maggioni reflects (in mail)[40] that there can be questions if a Subjectivist is more sophisticated than an Objectivist. This is an interesting argument and may have people aligned with either position especially when raised in the light of the paradigmatic debates. For the purpose of order in the scale, I argue in practical terms that the Subjectivist must come after the Objectivist on the scale because it is only after the epistemic

[39] It is important to consider the problem of convergence of meaning that may exist in a review of epistemic concepts. The activity involves interaction between the reviewers own concepts about knowledge, the range of concepts evidenced in the data, and the particular construct of epistemic beliefs that underpin the scale in use. Suffice to say that this is complex and requires more words than my document permits.

[40] Maggioni, L. (2013), CBKH (personal communication)

awareness of the knower can one appreciate the use of the methods of history to resolve the problem of subjective knowing. I have described a naive and sophisticated position for each category depending upon the integration of the elements. Thus, there is an element of zigzag seen in development of sophisticated beliefs.

A key development in the CBKH is the unraveling of elements of belief about knowledge in history as the Known, the Knower, and Ways of Knowing. This was seen as necessary as there is incongruence seen in the development of awareness of the known and the knower with the development of knowledge and skill in the use of heuristics used to evaluate sources and accounts in history.

3.36.2 Selecting the Instrument of Data Collection

In selecting an instrument, Duell & Schomer-Aiken further suggest that apart from issues of validity, reliability, etcetera it is important to consider the underlying theory. Questions to ask are: do I believe the theory, does the instrument truly assess the theory, how does this relate to my research questions Duell and Schommer-Aikins (2001). The instruments I selected were suitable as described as they assess epistemic beliefs, can be used for a pre-test and post-test measurement of change, and there is some variance in the construct measured to provide a broader picture. I do consider that the measures are sound theoretically and do assess the theory largely. Epistemic beliefs are difficult to conceptualise and more research will help to develop these. As pre-test and post-test measures of belief change, they have been useful. Patterns have emerged that could be interpreted although the BLTHQ and EBI would benefit by development.

While truth claims about epistemic beliefs are necessarily tentative, there is enough data in 3 measures over 166 response sheets equaling 498 measurements to give confidence in the measurement.

3.36.3 Categories of Beliefs about Knowledge in History (CBKH)

Progression Models of beliefs about history require careful planning to develop as Lee (2005, p.164) points out. The CBKH rubric was born from necessity to reliably assess a large number of responses. It is built upon arguments and insight in the literature, online discussions with Maggioni, patterns seen in the data, as well as what I learnt from teachers as they discussed their own and their pupils concerns in the course of teaching and learning history. The online discussions with Maggioni on epistemic issues were a significant resource and are cited in these pages. The CBKH is a modification and development from the original model of epistemic beliefs in history, the Copier, Borrower, and Criterialist developed by Maggioni et al. (2009a). The CBKH, Categories of Beliefs about Knowledge in History, is a set of 7 categories of epistemic beliefs pertaining to the domain of history used as a rubric to assess epistemic beliefs as read in responses. The CBKH was developed for the purpose of assessing responses to the Boscolo and Mason questions in my study. The development was an unplanned project that became necessary due to the difficulty of consistent assessment of 166 response sheets with complex epistemic statements. Developing descriptions and bringing order to layers of complexity is not an easy task neither can it ever be said to be entirely satisfactory. Necessity is the mother of invention and therefore, this scale was developed for use in this research. The support of Liliana Maggioni at this time was invaluable.

Epistemic beliefs have a complex nature and as Pintrich (2002) argues, there are diverse views about the nature of personal epistemologies and the role they play in learning and development. Epistemic beliefs are believed to be either domain general or domain specific (Hofer (2002, p.11). In the case of history, beliefs about knowledge in history are seen as specifically relating to or having what Rogers (1978, p.7) calls ‘a symbiotic relationship’ with, or underpinned by, practices of knowledge development by professional historians and the evaluation of history accounts by readers or users of history. Ashby cites Rogers (1978) to argue that only ‘knowing how’ ‘can give the right to be sure’ because it is the only

valid base for claims to ‘know that’. There is a requirement to differentiate between responses from pre-test to post-test according to the general level of standards of epistemic beliefs observable within that sample. This particular rubric has been developed for the specific purpose of assessment of history related epistemic beliefs of teachers as users of history. The context of this research is Pakistan and the sample is of O' Level history teachers in Pakistan. The Category definitions are here adjusted to the sample of teachers of history rather than to professional historians or expert users of history. The examples included for the Categories are therefore, from this sample.

I must point out that in most studies, the level of this expertise and the attendant concepts about knowledge are seen as low in novices although interventions are seen as making a difference. Wineburg (1991), VanSledright (2002), Maggioni et al. (2004), Maggioni et al. (2009a) Educational research focuses on the development of beliefs amongst children and teachers and these groups do not generally belong to the level of experts in historiography. Unraveling slightly, as I have done, the levels of expertise and their relationship to a theoretical standard of naive or sophisticated belief may help differentiate between such samples with more relevance.

The research uses a quasi-experimental methodology. The data and the scale were required to reliably assess epistemic stances of participants at pre-test and post-test in order to identify and describe change as a result of the interventions. Assessment of responses to the set of open ended BM Questions is focused on sorting and ordering epistemic beliefs about knowledge in history that may underpin the responses of participants. The 6 BM questions were open ended, therefore responses varied in content, structure and form. A holistic reading across questions appeared to help recognise underpinning concepts to sort and order these. I carried out reading in repeated iterations in order to recognize familiar words and phrases and identify a possible stance but found that assessment was not easy or consistent. I realized that a category scale was essential for reliable assessment due to the large number of response sheets (166).

The CBKH is developed for specific purposes from the original model of the Copier, Borrower, Criterialist categories. These were first employed with suggested transitions designed by Maggioni et al. (2004) but changed and modified according to the differences in responses as observed in the data. A study of the literature and online discussions with Liliana Maggioni with reference to the original categories has informed the development and this is described in detail in Chapter 4.

Grounding the categories was necessary to find a better fit. The sample was one of regular history teachers of O' Level History in Pakistan. It was felt that the set of categories, while closely following theoretical perspectives, needed to differentiate between samples of non-expert users of history. Unraveling of the various elements of belief was also seen as necessary in order to make it possible to assess responses as consistently as possible. If there is incongruence in development of concept and skill, it can be identified with this unraveled rubric and a score awarded. Finally, and importantly, the concerns expressed by the sample of teachers of history regarding the history accounts they taught and that were experienced in the classroom by pupils, needed to be reflected in the Categories. These concerns also have some support in the literature and are relevant to school contexts of learning and teaching history.

3.36.4 Qualitative Analysis of Responses

Assessment of responses is qualitative. 6 Open ended questions in the BM questionnaire require repeated reading and holistic assessment to assign each respondent to a single category amongst the 7 categories of beliefs about knowledge in history. The sample fits into 6 Categories, therefore the 7th is seen as hypothetical and not included in the assessment. Each respondent is assigned one descriptor Category according to best fit of the descriptions in the Categories, CBKH. Scores are also assigned with the descriptor. These scores are assigned to describe fine differences between performances, to infer possible change in beliefs due to the intervention. The scores are used for a quantitative analysis of stance and change in stance from pre-test to post-test.

Pre-post-test Questionnaires are assessed separately. The complete sample after cleaning of data is a total of 83 respondents. Both Control and Experimental Group members are assessed and descriptors and scores assigned in the same manner. Thus, a Qualitative Descriptor plus Quantitative Score is assigned. Each response is placed in a descriptive category and a single summative numerical score is assigned after calculation. Weightage was originally seen as necessary as often the change is more a matter of degree of emphasis, numbers of references. This is a finer difference than a change in Category. The weighting was meant to help differentiate between responses.

Note:

The Categories of Beliefs about Knowledge in History, CBKH, are listed in Table 3.1 below. Tables 3.5-3.8 provide an initial version of the CBKH with weighted scores to show development; the final version of the CBKH; explain the categories with evidence and elements of the known, knower, way of knowing as these are unraveled for the rubric; provide a range in responses; and finally some exemplars from responses to the Boscolo and Mason questions.

Table 3.1: The Categories of Beliefs about Knowledge in History –CBKH

Categories	Score
Naive Objectivist	1
Critical Objectivist	2
Naive Subjectivist	3
Critical Subjectivist	4
Naive Criterialist	5
Critical Criterialist	6
Sophisticated Integrator	**7**

3.36.5 Category Scale Options with Weightage

I planned and assessed responses to the weighted scale given below. However, Inter Rater Assessment suggested that weight increased the numbers of subjective decisions. This was then dropped to a single score on the basis of holistic assessment to the categories. This option is provided here.

Table 3.2: Degrees of Pre- Post Change

	Reduction	No Change	Weak Change Within Category	Weak Change From Category	Strong Change From Category
Score	-0.5	0	+0.5	1	1.5

Table 3.3: Degree of Consistency in Responses

	Conflicting Statements	Some Consistency	Consistency
Score	0	1	2

Table 3.4: First Category Scale with Weighted Scores

Serial no	Categories	Score	Weightage	Sum Score	Degree of Change - 0.5;0;+0.5+1;+1.5	Consistency Score 0-1-2	Final Score: Summation of Sum Score + Degree + Consistency
1S	Naive Objectivist	1	-0.5	0.5			
2S	Critical Objectivist	1	+1	2			
3S	Naive Subjectivist	2	-0.5	1.5			
4S	Critical Subjectivist	2	+1	3			
5S	Naive Criterialist	3	-0.5	2.5			
6S	Critical Criterialist	3	+1	4			
7S	**Sophisticated Integrator**	4	+1	5			

Note:

These were various options explored in the development of the Category Scale. These were discarded for the assessment as they tended to increase the numbers of subjective judgments making assessment less than consistent. They are presented for discussion only.

Table 3.5: The CBKH Categories and their Elements with Reference to Evidence in the Responses

Elements of Belief	Explanation	Range in Responses
Known	The word 'known' refers to an awareness of what is it that is known. Belief in history may range between history as a true reflection of the past, or a construction of the historian, or an academic judgment developed on the basis of evidence and reason.	The response demonstrates awareness of what is known. Responses range between: Considers what is known as objective knowledge of the past; continuous and complete knowledge; knowledge of the past as subjectively acquired; or an account developed from perspective to answer historical questions with the critical use of evidence employing key principles and heuristics. Responses to the Question, 'what is history' range from the isomorphic 'the past' and 'reflections of the past' to 'interpretation of the past' or 'study or investigation of the past'. Focus on facts as knowledge of past to be organised/ memorized/presented. Maggioni et al. (2004, p.188). 'perfect correspondence with the past and history'. References to history as 'isomorphic with past'.

Elements of Belief	Explanation	Range in Responses
Knower	The knower refers to an awareness of the mediation of the writer historian in the interpretation/construction of the historical narrative. The range is relative to a belief in true knowledge in history as objectively acquired; to a sense that mediation makes all knowledge in history suspect; to awareness that knowledge in history is possible as a construction of the writer historian using disciplinary methods based on evidence and reason.	The Response demonstrates awareness of the historian as mediating knowledge of the past; as user of methods of history to build an account or explanation. The Responses can range from: • The response shows a belief in history as 'falling from the sky' where the awareness of the historian is missing from the evaluation of the account. • A completely relativistic helplessness that knowledge in history is difficult or not possible; with complaints of bias and opinion. • An understanding of how history accounts are constructed by historians with careful, persistent inquiry, use of evidence and reason. The Knower implies: • Awareness of the knower in the construction of an account as a person with cultural, other, identity, perspective. • Reference to perspective, opinion, bias, position, interpretation, standpoint in responses.

Elements of Belief	Explanation	Range in Responses
Ways of Knowing	'Ways of knowing' refers to an awareness/knowledge/skill in the methods or procedures of the historian such as investigation, the use of heuristics, in the development of a historical narrative.	The Response demonstrates knowledge and skill of the processes of knowing history. Responses range from weak to growing, to expert knowledge of method. From one who considers truth to be a property of the source thus refers to true sources/false evidence; to reference to research, inquiry, vague/broad use of evidence and source to choose between explanations. At a higher level, responses show a range of knowledge and skill in the historical method along with an awareness of the constructed 'tentative', nature of knowledge in history. Reference to the following words will be looked for in the responses: Study, investigation, research, analysis, and particular heuristics used in the evaluation of sources in history. The respondent understands that the historian uses available sources to answer questions. This is a sophisticated awareness that source was not necessarily created to answer the historians questions in the future but can be used by the historian in a critical method to inquire into the past. At a higher level, awareness of using a pragmatic hermeneutics. VanSledright and Reddy (2014)

Elements of Belief	Explanation	Range in Responses
Integration	Integration of the objective and subjective components of knowing with way of knowing. Integration is seen to be an aspect of sophistication in belief where the three elements of knower, known, and ways of knowing, are integrated in the evaluation of an account. Knowledge in history is specific in laying down methods of acquiring knowledge in keeping with 'guild honed' principles of historiography.	Responses demonstrate no integration, some integration and integration of what is known with the mediating presence of the knower and the range of knowledge and skill. Refers to history as interpreted and the heuristics used to obtain more reliable accounts or to select between explanations. Refers to the nature of historical knowledge as based on the examination of evidence and argument. Attribution is seen as key by Wineburg (1991, p.83). The historian's presence is acknowledged in the construction of an account with the use of available resources and the historical process. The evidence is recognized as discontinuous, fragmentary, and not made-for-historical judgment but for a purpose of the time. As VanSledright et al. (2006, p.225) explains, history is an interpretive reduction. 'It is necessarily tentative, unstable, and less than the past.'

Table 3.6: The CBKH Categories and their Elements with Reference to Evidence in the Responses

Categories	Evidence in Responses
Naive Objectivist	
Evidence in responses to the Known	Refers to history as 'the past' 'reflection of the past', referential illusion not recognized as such. What is known is either true/untrue on the basis of facts. An objectivist position that does not question how an account is developed and the role of the knower. Focus on facts as knowledge of the past to be organized/memorized/presented.
Evidence in responses to the Knower	No reference to historian as conduit to the past. History as 'authorless'. No reference to history as constructed from standpoint or perspective.
Evidence in responses to Ways of Knowing	None or weak reference to method/evidence/source. No integration of knower and known.
Evidence in responses of Integration	No integration with Knower, Way of Knowing Knower, Way of Knowing, absent from concept of what is known.

Categories	Evidence in Responses
Critical Objectivist	
Evidence in responses to the Known	History as story of 'the past' or ‘knowledge of past’. Reference to facts and truth, reality.
Evidence in responses to the Knower	Historian as scribe, faithful conduit of past. Brief reference to knower in the sense of beginning reference to opinion or perspective. Beginning concerns about source. Evidence as missing, hard to get. Sources as direct testimony.
Evidence in responses to Ways of Knowing	Accumulation of facts and evidence. Some use of words like ‘look for, analysis, research’.
Evidence in responses of Integration	Beginning integration of knower with known as seen in reference to some words such as eyewitness, opinion, bias.
Naive Subjectivist	
Evidence in responses to the Known	Doubts about history as possible. Transition to doubt objectivity as adequate to knowing and refers to/ask questions about the knower. History as personal opinion, story, narrative. Statements about true or false evidence. Correct or incorrect knowledge.

Categories	Evidence in Responses
Evidence in responses to the Knower	Refers to historian. Aware of historian as human, mediating knowledge of the past from standpoint or perspective. Refers to the subjectivity of the historian as making a credible account impossible. Doubts about bias and opinion. Refers to personal opinion, views, perspective, and interpretation. References to nationality, religious perspective, personal opinion, own viewpoint or way of thinking.
Evidence in responses to Ways of Knowing	No reference to method as means to choose between different explanations. Aware of missing sources or sources hard to find. Refers to things unknown in history. Words like difficult, impossible to refer to decide between conflicting accounts or knowledge in history. No way of knowing the difference between accounts. History is seen as difficult. No way of knowing the difference between accounts. History is a matter of personal choice. None or little integration of knower, known, way of knowing. Link between evidence and source not understood as the action of the historian, the use of skill and knowledge of method to obtain evidence from the source in response to the question.

Categories	Evidence in Responses
Evidence in responses of Integration	No integration of method with knowing to overcome subjectivity. Increased focus on knower. Choice between conflicting accounts is seen as difficult, or not possible.
Critical Subjectivist	
Evidence in responses to the Knower	History still as difficult to know due to difference in views and perspective but some beginning reference to method.
Evidence in responses to the Known	Refers to historian/reader 'recognizes constructed nature of knowledge but the historical method is not deemed an effective tool to deal with problems of conflicting or missing evidence', Maggioni et al. (2004, p.188). Refers to subjectivity and difficulty in selecting between explanations.
Evidence in responses to Ways of Knowing	Beginning awareness of method. Some general, vague reference to research, analysis, finding out. Beginning reference to some method to resolve the issue.
Evidence in responses of Integration	Beginning of integration. Weak reference to method but without the use of words such as by or through, displaying a belief in method providing a way to judge between different explanations.

Categories	Evidence in Responses
Naive Criterialist	
Evidence in responses to the Known	History as possible despite difficulty and/or positionality. Confidence in account. May not refer to problems of account as constructed from perspective or position. History as product of research, investigation.
Evidence in responses to the Knower	Refers to knower as historian, researcher, or scientist. Subjectivity may be overcome with use of evidence and argument. Standpoint or position of historian may not be questioned. Method alone may be seen as adequate to justify account.
Evidence in responses to Ways of Knowing	Refers to method. Reverts to/Recognises the effective use of method to overcome the problem of subjective knowing. Refers to analysis of reliability and authenticity in reference to source. Includes some direct reference to heuristics used in judgment between explanations in history. Some reference to context.

Categories	Evidence in Responses
Evidence in responses of Integration	No acknowledged integration of subjective and objective dimensions of knowing. More informed reference to research/evidence/source/heuristics that may help to evaluate an account. History account makes little or no reference to historian or to perspective or standpoint. No reference to knowledge as tentative or historians personal point of view. No discussion on opposing views.
Critical Criterialist	
Evidence in responses to the Known	What is known is a product of research and historical inquiry. A product of judgment based on evidence and argument. The subjective presence of the historian in the construction of the account *may* be referred to.
Evidence in responses to the Knower	Reverts to some awareness of the historian as knower, conduit to the past. May recognise the constructed nature of the historical account and the historian as user of processes of history. May refer to interpretation and perspective again.

Categories	Evidence in Responses
Evidence in responses to Ways of Knowing	Importance given to knowledge and skill in use of heuristics. Refers to heuristics *but not necessarily with complete expertise.* Several informed reference to heuristics in making an account reliable than as Naive Criterialist. Refers to the purpose for which the source was created. *May refer to the questions asked by the historian of the evidence.* More reference to reliability, authenticity*, (Authenticity is a commonly used term amongst teachers in Pakistan for some reason. I understand that authenticity may not imply reliability in the respondents language use and may refer more to truth and reality as in the 'real' artifact. Refers to use of source as evidence. *Use of source to answer questions the source was not created to answer not directly seen in my data.* 'Aware of disciplinary tools to transform primary sources into evidence' Maggioni et al. (2009a, p.194)
Evidence in responses of Integration	Integration of knower and known underpins the view of knowledge. What is known can be an integration of method with the subjective presence of the historian and the philosophy of the historian. Aware of presence of historical investigator who can ask questions that sources were not specifically designed to answer. Lee and Shemilt (2003) (Weak evidence in my data to this) Maggioni et al. (2009a, p.195) Ability of historical thinkers to use the disciplinary tools and criteria for historical inquiry and to focus on a multiplicity of particulars without losing the capacity to perceive a broader view.

Note:

The Sophisticated Integrator stance is hypothetical, not counted in analysis.

No participant amongst my respondents was seen as fitting at this level mainly due to a weak knowledge of heuristics although integration is evident in places. References to the acknowledged presence of the historian was brought up and debated quite often in the workshops by the participants but perhaps the questionnaire does not provide a slot to add that element. There is evidence of this in the video records and field notes. The key drawback is a lower level of knowledge and skill in the evaluation of accounts using the methods of history.

Table 3.7: 'Sophisticated Integrator' of the Subjective and Objective Dimensions of Knowing

Category	Evidence in Responses
Sophisticated Integrator	
Evidence in responses to the Knower	Recognises the historian as scientist and mediator, subjective, human, conduit of the past. Historian with cultural baggage, personal theories, and member of community of scholars. Possessing expert skill and knowledge of method. Questions the lack of an acknowledged presence of the historian in the text, VanSledright (2014).
Evidence in responses to the Known	Recognises history as an authored narrative, as tentative judgment, constructed from traces, discontinuous fragments, residua, relics, of the past. Recognises/refers to the interpretive, constructed, nature of historical knowledge. Recognises/refers to the tentative nature of historical knowledge. Recognises/refers to/questions, the *acknowledged* role and presence of the subject in constructing an account. Recognises the history account as best explanation under the circumstances of weak, conflicting evidence, and gaps in evidence. Recognises the history account as developing and open to challenge, change and evolution.

Category	Evidence in Responses
Evidence in responses to Ways of Knowing	Sophisticated integration of known, knower, way of knowing. History as constructed through careful, transparent, process and coherence with body of work, in acknowledgement of subjectivity of the historian. Use of method with knowledge and skill. Has knowledge, skill and belief in the historical method. Recognises the role of questions in evaluating a source. Can evaluate an account on the basis of question. Can identify/use key heuristics for evaluating a source to find evidence. Deep understanding of context. Use of empathy. 'the search for the best explanation through the patient weaving together of the best evidence and the best argument available.' Maggioni et al. (2004, p.187) Pragmatic heuristics see VanSledright and Reddy (2014, p.187). Pragmatic or other transparent, open lens or epistemology.
Evidence in responses of Integration	Integration of subject and object with the way of knowing as medium of integration. Accounts are referenced and personal theories of historian are stated and open to argument. Makes transparent, or requires transparency of the role of the author in constructing an account. Is critical of missing authors and the referential illusion.

3.36.6 Selection of a Range of Responses for Each Question

The range of response excerpts is ordered here broadly according to Category levels. The integration of the objective and subjective dimensions is a criterion for differentiating between responses. From a stance of knowledge as objective truth to an awareness of the subjective presence of the historian to a greater integration of the use of evidence and method with reference to the mediation and positionality of the knower is the general order. Nuances of meaning in the responses make clear ordering difficult but interesting. The score is awarded in a holistic reading of responses to all 6 questions. A response to a single question is not adequate for judgment into a separate category.

(The Exemplars are edited for grammar or sentence structure unless meaning was affected.)

Table 3.8: Exemplars from Responses to Boscolo and Mason Questions

Question	Evidence in Responses
Question 1 **What is history?**	The past. Reflection of the past. History is what had happened in past. Events of the past, which can sometime effect our present and future. Story of past. An interesting story that connects a person's present with past. Historian's interpretation of past events. Study of Past. Insight of past. It is method of obtaining truth, opinion and find out the causes of the event and fact through effect. History is the interpretation of past events based on valid evidence.
Question 2 **How do people who write history know about the past they write about?**	Books and Journals, newspapers, internet, films, study previous books, reading resources. These responses indicate a teacher-focus rather than one referring to sources of knowledge. *References to sources as objects* Collect facts, In light of facts and figures. By collecting information. By collecting data. Eyewitnesses, past experiences, listening to stories, detailed accounts, biographies.

Question	Evidence in Responses
	Historical buildings, official documents, theories and manuscripts, chronicles, poetry, painting, excavations. Archeological sites, primary, secondary sources. Different resources, pottery, ceramics, currency, word of mouth. Through evidences provided, through sources. *References to process.* By checking, doing research, investigating. Through different sources i.e. gathering data, interpreting information, collecting evidences. By feeling into the past. *Evidence of integration.* Collect evidence along with personal opinion. Constructing. By deciding which is best depending on the reliability of the source.

Question	Evidence in Responses
Question 3 **What problems can historians have when they try to understand what happened in the past?**	Fabrication. Lies. Bias. Definition of right or wrong is different. Sources are not authentic. *Reference to problems of sources and methods.* Sources can be biased. Misrepresentation of facts, distortion of facts. Unreliability of sources. Sources are false, true, or untrue. Problem is only of finding and collecting data. Has to totally depend on sources. Lack of right evidences. Evidence is missing, hard to get. Adopting methodologies which are not cost effective or time effective. Contradictions in data. Contradicting sources provided by different people. Contradictory accounts. Contradiction in research of the past, official reports. *References to historian and perspective.* All sources may be to present a certain view. Authenticity of sources, perception of author. Facts will remain facts but stories can be interpreted in different ways. Weak sources, incomplete knowledge, a historians cultural baggage, Interpreted wrong or a one side picture is shown. Each historian has different outlook.

Question	Evidence in Responses
	People look at the event from different views. They have their own interpretation and opinion about the events. To judge between accounts. To find a peculiar truth.
Question 4 **Is it possible to explain what happened in different ways?**	In most responses the answer was a Yes. Some elaboration was provided.
Question 5 **Why?**	Often not responded to. Some sheets say 'already answered'.
Question 6 **If there are two different explanations, how is it possible to understand which is better?**	Responses indicating helplessness in the face of conflict or difficulty in access to evidence. Can't think of an answer. Difficult to answer. Beginning of integration. Difficult to prove but yes can be possible by gathering facts and the dates and the events. References to subjective judgment as basis of choice between accounts. The one you feel according to the situation is better. People have different perspectives. Explanation near to your way of thinking.

Question	Evidence in Responses
	The explanation which is authentic and closer to our personal opinion. That depends upon the person's ability, mode of judgment. By studying all available interpretations and evaluating which is better. Search for disinterested sources. References to method, heuristics as ways to decide between explanations. Different way of analysing. Through using different sources, then analyzing to check its authenticity. Whichever justifies with evidence and proves it with some kind of source will be a better answer. Investigate reliable evidence. How good the source is for those explanations. Through various primary and secondary sources. Through finding a number of different resources that are reliable. Number of sources and evidence. Evidence is like fish on a fish mongers slab. Some are rejected, others are kept. Better one is followed by the primary resources, reliable evidence, or authenticity. Because the sources may be provided for different purposes e.g. official documents or to guide the people to the actual situation. The number of evidences must be considered for a certain view. It also matters who writes the view whether the explanation is written by the person who was a part of the event or had influence. So evidence alone can be helpful. Historian acts as lawyer to present interpretation and judge to check validity.

Question	Evidence in Responses
	Because history is my perspective in the light of certain reliable sources which still satisfy my set of ideas and requirements. How valid is it? Who wrote and why he wrote!

3.37 Validation of Assessment

3.37.1 Assessment of Reliability of Marking of Responses to the Boscolo and Mason Questions

The Boscolo and Mason questions are open ended and the narrative responses are open to interpretation. The large number of response sheets (166), required a rubric for consistent and reliable assessment. The CBKH was developed and is described in the Chapter on Research Methodology and argued on the basis of theory in the literature review.

I marked sheets myself and checked corresponding scores in early marking and calculated the reliability of assessment. In order to make the process rigorous, I developed the rubric, the CBKH. This was a time consuming process taking newly two years and was supported by Dr Maggioni as described. With each assessment of responses to the developing rubric, I revised the categories and identified exemplars to improve the assessment. These revisions are provided in the document in Section 3.1 onwards. I had, for example, tried weighting the scale but realised this added to numbers of subjective decisions and therefore, changed the framework.

A rigorous process of first, Inter Rater Assessment, followed by Internal Evaluation, was conducted. The process informed the development of the CBKH as well as assumptions of reliability of the assessment. It is concluded that rating of responses requires trained raters with the requisite knowledge and skills. A sound grounding in the principles and methods of history as well as sophisticated awareness of epistemic beliefs in history would be necessary for reliable assessment of epistemic beliefs about knowledge in history even when using the CBKH Categories.

The Internal Evaluation was conducted used blind marking after obtaining numbers generated by a Random Number Generator from the complete set of 166 response sheets. Internal evaluation was conducted three times with improved consistency. These results are reported in detail in Tables at Appendix A7 (A7-1,

A7-2, A7-3 and A7-4). Summarized results of Table A7-4 are also reproduced below.

Table A7-4: Comparison of Three Internal Evaluations

Internal Evaluation	Percentage of Match	Percentage of Mismatch
1st Internal Evaluation	65%	35%
2nd Internal Evaluation	53%	47%
3rd Internal Evaluation	82%	17%

The improved categories and the use of exemplars made assessment more reliable as Table A7-4 shows. 82% reliability was taken as adequate considering the nature of the construct of epistemic beliefs.

3.37.2 Inter Rater Agreement for Validation of Category Table Inter Rater Exercise

To test The Categories of Beliefs about Knowledge in History as a rubric to assess and order epistemic beliefs about knowledge in history:

To test reliability of assessment of responses through an inter rater agreement exercise.

3.37.3 The Raters

In order to find raters for this rather specific and somewhat challenging rating, I visited the History Department in a University in Pakistan. The only people available were 3 students of M Phil History. One had completed his study, one was in a research phase and the third was still completing course work.

Upon discussion and trial, I realized that their knowledge of the historical methods and principles of history did not match the requirement. Therefore, I undertook to train them over 4 sessions. This is described under Process.

3.37.4 Selecting a Sample for Inter Rater Analysis

A sample of responses was selected for Inter Rater Analysis. A simple random sample developed using a Google based programme, 'Random Number Generator', was employed to select a sample for Inter Rater Analysis. 60 % of the total was worked out and the programme was set to generate the sample. Duplicate numbers were allowed. Response sheets were copied for Raters, and pre-test and post-test sheets were mixed and not identifiable. The number 51 was selected, as 60% of the total was selected for Inter Rater Agreement. 60% of 84 was 50.4. This was rounded off to a whole number 51.

I decided to use duplicate numbers, as the total, being small, and the number of Raters could carry out the assessment separately; duplication would add strength to the rating given. Each number would have an equal chance of

being selected. This would also help me see if raters marked the same script the same or differently. An issue was that all response sheets were not marked by the Raters. This was limited to the time the raters could give. In all, 4 sessions of training and discussion including moderation of samples were all that were possible.

3.37.5 Process

The Raters were first assessed for the concepts and skills required for the task. Next, they were trained. The Inter Rater exercise was held after training. I used the same history material of my workshops, to help build concepts, in order to keep contexts comparable. After the training, they marked some responses each, and discussions were held to compare scores. Raters were allowed to discuss and question each other. I, the researcher, participated in the sessions. There was free dialogue and argument which was productive in developing understanding. After the training, they were given a post-test. The results are provided later in this Chapter.

On the 4th day, a revision session discussing the categories with examples was held again. This was followed by the final marking exercise using the complete, randomly selected sample. Copies of the sheets were distributed and the Raters marked these. They were given the Category Table in its original form plus a simplified version for ease in analysis. Data of marking was recorded on Excel and a comparison was made. I sat in the session and discussed any sheet they wished to talk about.

3.37.6 Tasks for Raters

- Read responses to all 6 questions holistically.
- Use the given CBKH rubric to assess responses.
- Assign a single category to each respondent with a single mark.

3.37.7 Questions for the Inter Rater Moderation

- Are responses marked consistently by individual raters?
- How similar or different is the marking of different raters?
- How similar or different is the marking of raters and researcher?
- Do raters perceive a change in responses?
- What is the direction of the change?
- What is the degree of the change?
- How does the marking of change by raters compare with that of the researcher?

3.37.8 Criteria for Inter Rater Marking

- A difference in a category awarded is seen as a failure. A difference between scores that reflects a miss between a subjective or objective position is seen as a critical failure.
- A difference in degree within the same category awarded on the basis of a use of method is seen as a pass.
- A match of scores is seen as a strong pass.

3.37.9 Discussion on the Process of Inter Rater Agreement for Validation of the Category Table

This validation required raters to have a sound understanding of knowledge in history. To be able to use the Categories to assess epistemic beliefs about knowledge in history ideally, the raters needed to themselves hold beliefs concurrent with the system, at the highest levels of sophistication. Secondly, Raters, to make a comparable assessment of responses, needed to be at similar levels of knowledge and skill amongst themselves.

3.37.10 Concepts and Skills Required of the Raters

- Concepts about sources as relics of the past. Awareness that such relics become weak or reliable evidence depending upon the questions posed by the historian.
- Concepts about the difference between sources and evidence.
- Knowledge about types of sources.
- Understanding of historical investigation.
- Knowledge of heuristics such as asking questions of the purpose for which the source was created, the author of the source, and to ask questions for which the source was not created to answer.
- Concepts about the constructed nature of knowledge in history.

These are the concepts the categories are based upon, and raters needed to hold in order to consistently, dependably, rate responses to questions about knowledge in history. The raters often used terms such as perspective, opinion, heuristics, but did not themselves demonstrate sophisticated concepts of knowledge in history. Scores obtained by the students in a pre-test are given in Table 3.9 below.

Table 3.9: Pre-Test and Post-Test Assessment of Raters' Concepts about Knowledge in History (Instrument: responses to the Boscolo & Mason questions)

Code FM	FM has a sensitive understanding of complexity of history but also says 'History is a complete description that bridges over past and present'. She refers to perspective and to 'personal biasness of author', 'nationalist perspective', and 'different perspective'. She refers to method and 'reliable, unreliable sources'.	4	FM grows in sophistication. Refers to knowing through research, evaluating, synthesising primary, secondary sources, 'extracting truth through relying on reliable sources'. Refers to heuristics, by knowing about the writer, his/her purpose, and how that historian used that source. She also, however, attributes weak methods to the historian, 'not bothering to know...' Refers to a nationalist and liberal perspective. She refers to the questions posed by the historian and to resolving differences in explanations by using heuristics, questioning the agenda/ purpose of writer.	5
MM	MM refers to history as 'simply the past events' and examining 'facts and figures'. Conversely MM refers to primary and secondary sources 'problems with availability of sources', and 'biased	3	In the post-test, MM is brief. He refers to history as the study of the past, through different primary, secondary sources. He again refers to 'biased approach'. He refers to analysing two different explanations through 'further research i.e. background of authors, sources used, biased approaches'. I infer from	5

	approaches', 'personal details of the author also helped to judge the better explanation'		the above that he is referring, albeit weakly, to heuristics as well as perspective.	
MT	MT quotes Marx and Hagel, refers to philosophy of history. Refers to 'sources or evidences as reflecting the past'. Says quality of argument can help tell which explanation is more appropriate. Refers to multiple angles but not subjectivity. Says it is not easy to refute, face problems. Does not refer to heuristics.	2	MT refers to method, analysing and evaluating the authenticity and reliability of sources via cross checking sources. Says event can be explained by reference to perspectives and opinions; Vaguely refers to heuristic, 'by cross checking different opinions, for an integrated generation of a new approach/opinion'. The last refers to common research practice in Pakistan.	4

3.38 Comments on the Exercise

The Rater Training Exercise was successful to a limited degree in developing the Raters' epistemic beliefs in knowledge in history. The second rating was perhaps the experience that best helped develop ideas. The Inter Rater marking of the Questionnaires did not demonstrate a strong, sound and consistent growth in the Raters, but somewhat supported both the Categories as a workable instrument as well as the rating I had given to the responses.

The data is qualitative as is the method of analysis. The target is epistemic beliefs about knowledge in history. Epistemic beliefs, in spite of the difficulty in accessing them, are being assessed in various ways around the world. This is due

to the importance of the beliefs for learning and development in various subject areas.

The Raters had little prior knowledge of epistemic beliefs, this being a first time or second time that respondents had ever attempted such questions. An automated selection of words to assign responses to categories did not seem to be appropriate. A holistic judgment, an inference, is needed to be made, therefore, although I had my doubts about the success of an Inter Rater agreement exercise for this data, I have attempted it. I did not expect a reliability quotient to be significant but I just wanted to know how usable were the categories by others, and how far could the analysis be replicated.

3.39 Issues

As described, the prior knowledge of the Raters was weak. Training made a difference in their familiarity with the responses and the category table. However, this was not adequate for a sound, consistent, reliable Inter Rater agreement according to the questions framed above.

There was, however, a reasonable degree of development in their use of the instrument, some consistency in the marking, and most importantly, a growth in sophistication towards a more consistent rating of category if not in degrees between it.

This was not adequate in rating the instrument as reliable. Judgments are a holistic inference. There was, however, reliable and consistent use of the instrument enough to suggest that with more sophisticated expert raters, the category table could provide a reliable judgment. An internal evaluation was next carried out to improve the reliability of the rubric and assess the consistency of my own marking.

Table 3.10: Inter Rater Assessment of Responses Comparison of Scores by Researcher and 3 Raters

Ser	Code	M	T	Mo	Researcher	Code	M	T	Mo	Researcher
1.	CR5aBM				2	CR5bBM				2
2.	CR8aBM	2			3	CR8bBM	5			4
3.	CR10aBM			3	3	CR10bBM	2			4
4.	ER13aBM	4			4	ER13bBM	5			4
5.	ER14aBM				3	ER14bBM				3
6.	ER15aBM	4			2	ER15bBM	4			5
7.	ET2aBM		4		4	ET2bBM		3		5
8.	ET3aBM		4		2	ET3bBM		3		5
9.	ET6aBM		4		3	ET6bBM		4		3
10.	CT9aBM		3		1	CT9bBM		3		1
11.	CT10aBM		4		3	CT10bBM		4		4
12.	EJ1aBM	3			4	EJ1bBM	3			4
13.	EJ2aBM	5	4		5	EJ2bBM	3	3		5
14.	EJ4aBM	2			3	EJ4bBM	6			4
15.	EJ5aBM	3			3	EJ5bBM	2			4
16.	EJ6aBM	2			3	EJ6bBM	4			4
17.	EJ8aBM	2			2	EJ8bBM	3		2	4
18.	EJ14aBM	3			3	EJ14bBM	4			4
19.	CJ16aBM	3			3	CJ16bBM	3			3
20.	CJ18aBM	3			3	CJ18bBM				1
21.	CJ21aBM		4		5	CJ21bBM		4		5
22.	CJ22aBM		4		1	CJ22bBM		3		1
23.	EB3aBM		3		2	EB3bBM		2		2
24.	EB4aBM		3		4	EB4bBM		3		4
25.	EB5aBM		3		3	EB5bBM		2		1
26.	EB6aBM		3		1	EB6bBM		3		2
27.	EB8aBM			3	2	EB8bBM			4	2

Ser	Code	M	T	Mo	Researcher	Code	M	T	Mo	Researcher
28.	CB13aBM			2	4	CB13bBM			2	4
29.	CB15aBM			2	2	CB15bBM			3	2
30.	CB17aBM			3	1	CB17bBM			3	2
31.	CB21aBM			2	1	CB21bBM			2	2
32.	CP2aBM				4	CP2bBM				4
33.	CP6aBM			3	3	CP6bBM			3	3
34.	CP7aBM			3	2	CP7bBM			3	2
35.	CP8aBM			3	2	CP8bBM			3	4
36.	EP9aBM			3	3	EP9bBM			3	4
37.	EP10aBM			2	1	EP10bBM			2	1
38.	EP11aBM			2	1	EP11bBM			2	1
39.	EP12aBM			4	4	EP12bBM			3	4
40.	EP13aBM			2	1	EP13bBM			3	2

Response sheets assessed by Participants in Inter Rater Exercise: 73

Response sheets assessed by Researcher in Inter Rater Exercise: 80

Matched Scores: 32/73

Category Difference: 41/73

Percentage of Matched Scores: 44 %

Percentage Difference of Mismatch: 56%

3.40 Summary of Chapter

The research methodology employed in this research has several crucial elements that are described and discussed in the Chapter on Research Methodology. Aim and objectives of the research, objectives of design, data collection and analysis, key features, are listed in an easy to read, ready reference. Importantly, the background of the study, procedures followed, and features such as constraints and opportunities are described in detail. Information about the two interventions is provided and a key section details how a perception of agency is

operationalised and provided in the experimental groups. Quasi experimental theory demands that threats to inference be identified in advance and then steps be taken to circumvent effects. This is reported in the Sections 3.30 to 3.34 in transparent detail.

The development of a rubric to assess the 166 response sheets of the Boscolo and Mason questions became necessary. The development and key features of the rubric, the CBKH categories, is presented. Stages of the development are displayed in tables, and key features detailed that make the instrument available for use. Inter rater and internal evaluation carried out to assess reliability is reported in the end.

Paradigms, mixed methods and truth claims are discussed at length in the Review of Literature. I have concluded that ‘what works’ is adequate as a pragmatic philosophy, if it is known how it is known, and to what parameters it can be held to apply. Detailed and transparent reporting of each aspect of research methodology is a step in this direction.

CHAPTER 4

RESULTS AND DISCUSSION

'But after your workshop, aa the information I gave that aa through authentic references, references should be authentic and you analyse... that aa how you analyse the themes, these two things aaaa will help you more to differentiate the two examples or two explanations ... the difference between (yes yes) the twooo... the difference between the two.. if there are two explanations it ***is*** possible to understand which is better.'

- Participant in focus group comparing differences in his responses to Boscolo & Mason questions at post-test.

4.1 Introduction

Chapter Four presents analysis of the data. The data of each instrument is analysed separately and findings are discussed. References are made to compare findings of the three instruments where required. Results from tests are presented and discussed briefly under the Tables and in the Sections for ease in reference. Appendices contain the required tests.

Briefly, the research is conducted in 5 quasi-experimental small studies. Each study has a Control and an Experimental Group. The sample is non-random with a total of 83 participants, unevenly divided, as found, into Control and Experimental Groups. Meta-analysis is conducted of the overall combined data of the 5 studies in order to raise the numbers and obtain an overview. Independent studies are analysed and the results reported where relevant and possible especially for BM data. Three measures are used, two Likert Scales, the BLTHQ and EBI. One questionnaire with 6 open ended questions provides qualitative data, which is assessed, and a single holistic score is awarded to each participant. Change in beliefs is examined from pre-test to post-test to answer study questions.

Data from a focus group, audio records of sessions, and field notes, add to qualitative descriptions where possible. Tests and numerical analysis of the data include, Chi Square, Cramer's V, Effect Sizes, Odds Ratios, Wilcoxen Rank Sum Test, and the Mann Whitney for small studies. An estimate of the Difference of the Difference is carried out and reported. Factor Analysis of both the BLTHQ data and that from the EBI is conducted and reported.

The research objective is to explore a relationship between a perception of agency in task based group discourse and change in concepts. This translates into the question, if there is a difference in change of epistemic concepts from pre-test to post-test between Control and Experimental Groups, and if there is evidence of more change in the Experimental Group. Causal relations are difficult to claim especially in low powered studies as discussed by Shadish et al. (2008). However, the Experimental Group, having received the special treatment, the results are expected to be different in this group as compared to the Control Group. Intervention one was a rich educational experience that was provided to both Control and Experimental Groups; therefore, some change in epistemic beliefs was expected to take place in both groups. The added impetus to change in the form of a perception of agency provided to the Experimental Group alone is expected to have made a difference to conceptual change if the theory is sound.

To reiterate, concepts are found to be difficult to change even in task based discourse where contexts, actions, and words are present. An enhanced perception of agency relating to change is examined in this study. If the relationship holds, the experimental groups should display a better performance at post-test.

I do not expect significant difference as the intervention of a perception of agency is a psychological construct studied in educational contexts. This relationship has not been researched before, in my knowledge, for me to take a reference for level of significance. Quasi-experiments differ from true experiments in the extent to which causal inference can be made. In the light of concerns about non-random selection and assignment, the role of the researcher in the conduct of the interventional workshops among other threats, an alternative hypothesis is proposed.

Reliability was tested for BLTHQ data using Cronbach's alpha and was found to be 0.752 at post-test for the Control Group and 0.639 in the Experimental Group at post-test. I have some doubts about using the Cronbach alpha for the BLTHQ as there are 3 sets of items according to stance in the instrument, therefore internal consistency may be difficult to assume. Apart from this, Factor Analysis was conducted and the results are reported. Results appear to have some common features with the author's findings. Reliability of each factor was tested and found to be good i.e. in the range of 0.679 to 0.778 in the Control Group and 0.602 and 0.806 in the Experimental Group.

Effect sizes in educational research are explored and a moderate effect of 0.500 Cohen et al. (2007) is considered taking guidance from quoted Hattie rankings of effect size in educational research. The rankings given for motivation 0.48, reducing anxiety 0.4, a disposition to learn 0.61, affective attributes, 0.24, are considered for a rough estimate as these relate to similar psychological constructs in evidence based research. Geoff Petty on Cambridge teacherstoolbox.co.uk suggests that an effect size of 0.4 is above average for education research. Instead of a null hypothesis, I frame an alternative hypothesis that there is a difference in change between Control and Experimental Groups and look for evidence to support this.

The analysis focuses on identifying the stance of respondents and change in stance from pre-test to post-test in both Control and Experimental Groups. Results and discussion are concurrent and juxtaposed in order to facilitate understanding.

The BM questions are more theoretical and general relating to epistemic concerns while the BLTHQ statements are set in the context of pedagogical practices of history teachers. The EBI explores general epistemic beliefs rather than beliefs in the domain of history. The nature and contexts of each instrument being different, I examine change as observed in each instrument separately expecting a different reading. My study compares the effects of the intervention, a perception of agency on change in concepts in task based group discourse in teacher education classrooms. The alternative hypothesis states that there is a

difference in change of concepts in Control and Experimental Groups. The stance itself that respondents assume will be interesting but not crucial to the support of the hypothesis. I am, however, mindful of the limitations of the research, and these are described.

The BM questions are more theoretical and general relating to epistemic concerns while the BLTHQ statements are set in the context of pedagogical practices of history teachers. The EBI explores general epistemic beliefs rather than beliefs in the domain of history. The nature and contexts of each instrument being different, I examine change as observed in each instrument separately expecting a different reading. My study compares the effects of the intervention, a perception of agency, on change in concepts in task based group discourse in teacher education classrooms. The alternative hypothesis states that there is a difference in change of concepts in Control and Experimental Groups. The stance itself that respondents assume will be interesting but not crucial to the support of the hypothesis. I am, however, mindful of the limitations of the research, and these are described. Results and discussion are concurrent and juxtaposed in order to facilitate understanding.

4.2 Brief Outline of Objectives and Analysis Decisions in the Context of the Research

The authors, Maggioni et al. (2009a), argue that the items of the BLTHQ are reasonable proxies of the epistemic beliefs named the Copier, Borrower, and Criterialist. Thus, agreement with the items for a stance can be taken to denote the epistemic stance of the respondent for the purpose of the analysis.

I am conscious of the fact that the numbers of participants were limited, assignment and selection was non-random, control was limited, this being an experiment in the field. Many threats in the quasi experiment (as argued by Cook and Campbell, Shadish Cook and Campbell and discussed at length in this document), were controlled but there were limitations to this. I, as practitioner/researcher, conducted the intervention workshops and made

decisions. Groups were self-selected or selected by means of which I had no control as is the nature of educational workshops in schools. I have not offered a null hypothesis at outset and only look for enough evidence to support an alternative hypothesis. I also want to explore ways that this relationship can be researched by my colleagues in the field. The instruments provide rich and varied data although the sample has its limitations.

Normal distribution is not the norm in the 5 study groups. Control and Experimental Groups are also not built on random assignment but are self-assigned according to location, preference, etcetera. Age, language capability, etcetera are other variables. Apart from demographic differences, there are many other variations within the sample. The study groups have a range of differences in numbers in the groups as well as in demographics, such as the number of women is far larger than men. Prior subject qualifications in terms of study towards a Masters in History, or Pakistan Studies with a history component, are key differences. All are teachers of history but it turned out that in one small group there were some teachers of history from a junior level where they taught World History and not Pakistan History. Attempts are made to use tests suited for the data. However, in places, tests that are required but that do not strictly meet statistical test criteria are used anyway, and the interpretation is left to readers. (For example, in the Factor Analysis, the KMO Bartlett test of Sphericity result is ignored and Factor Analysis is carried out. I have been creative in trying to observe a pattern in cross tables of the 21 items by arranging test results in Table 4.16 for analysis. Since this became useful I repeated the format for other analysis).

In this sense, the context of the research would be familiar to other practitioners in the field of teacher education. The workshops, where the interventions were conducted, were like all other such teacher education experiences that the teachers were used to attending. The research conclusions may fall short of being highly reliable but may be valid in the sense that they are practitioner experiments in the field.

I used the BLTHQ Likert Scale instrument as 1 of a set of 3 measures to explore the change in beliefs of the participants in my research. The scale was a

6 point measure. At the time of conducting the research workshops, I altered the instrument to a 5 point scale adding a neutral in place of a forced choice, keeping all items intact and adding a brief glossary to explain some words used. This is explained in the Section on Measures. The total data is of two groups, Control and Experimental, with 40 and 43 participants, a total of 83 teachers of O' Level History. The research was conducted in small groups in 3 major cities of Pakistan. The numbers above are after data cleaning.

A meta-analysis was conducted first of the overall combined data followed by analysis of the 5 small studies in the three cities. Each study had a Control and Experimental Group. The 5 groups are Rose-A, Tulip-B, Jasmine-C, Bougainvillea-D and Pansy-E. A pre-test and post-test of both Control and Experimental Groups is assessed and the two groups are compared to explore the change in responses due to the two Interventions. To repeat, Intervention A was provided to all and therefore, some change is expected in all participants' responses. Intervention B was provided to the Experimental Group alone, therefore a difference is examined in response to the research questions. An important relationship between Interventions 1 and 2 with change is discussed at the conclusion and within the discussions as required. The Research Map (Figure 3.4) is provided which helps to explain this.

4.2.1 Missing Data

Missing data is due initially to drop outs after the first day when the pre-test was held or late entries after the first day. In some cases, participants attended the day but did not wait to complete the test and submit it. In other cases all three instruments were not completed. All of these participants have been discounted. Those measures of respondents who had submitted all three instruments have been retained even if some items were not marked or were hard to read. The few missing responses after screening of the data in each item are replaced by nearby median points as my data is ordinal and the median point is seen as appropriate. Data was also deleted of those few participants who attended the workshop but

were not teachers of O' Level History from Classes 9 onwards as they would have been teaching World History and not Pakistan History.

4.3 Analysis of Data from Responses to the Boscolo & Mason Questionnaire

4.3.1 Brief Introduction to the Analysis of Data from Responses to the Boscolo & Mason Questionnaire using the CBKH Category Scale

The Boscolo & Mason questions, henceforth referred to as the BM questions, are the first of three instruments used in this research. The responses are in a free narrative form and require repeated reading and interpretation. After analysis of the response of each individual to the BM questions a qualitative descriptor is awarded on the CBKH Categories.

The Categories range from 1 for Naive Objectivist, 2 for Critical Objectivist, 3 for Naive Subjectivist, 4 for Critical Subjectivist, 5 for Naive Criterialist and 6 for Critical Criterialist. The scores transform the qualitative descriptors into numbers and these scores are then analyzed in quantitative terms.

The main focus of my analysis in response to study questions is to examine the change in position of respondents' stances in pre-test and post-test in both sets of data, Control and Experimental. This will be the basis to support a possible relationship between a perception of agency and change in epistemic beliefs, which is the focus of my research. A secondary focus is upon the effect of an experience of a range of suitable examples upon change. This was an interventional experience provided to both groups, therefore change should be evident in both sets of data. Finally, the question if this relationship of a perception of agency on change in concepts can be researched will be discussed.

4.3.2 Questions for Analysis

- Is a change in epistemic beliefs observed in responses of participants from entry pre-test to post-test after intervention in both Control and Experimental Groups?

- Is there a difference observed in change of epistemic beliefs between Control and Experimental Groups in terms of numbers, nature, or direction of change?
- Is change from pre-test to post-test and between the Control and Experimental Groups supported in all three measures?. What are the differences in change observed in the three measures, the Boscolo & Mason questions, The Beliefs about Learning and Teaching History Questionnaire, and the Epistemic Beliefs Inventory?
- How reliable and significant is this observation of change, and difference in change, between Control and Experimental Groups?

An important reminder is needed here that some change is expected in both groups, Control, Experimental, due to Intervention 1 as described. The main experimental intervention, Intervention 2 is expected to effect change in the Experimental Group only. A difference in change between groups will thus be seen to relate to the difference in experience provided only to the Experimental Group.

The analysis of data is conducted at various levels and categories. An intense search is conducted in a meta-analysis of combined data from all five studies followed by an analysis of each study. The data is combined primarily to increase the numbers for statistics and each group is also studied separately to find the difference in groups and related interesting information. The data, as mentioned above, is in a single score per respondent for each test, pre and post, on the scale CBKH.

It is important to note here that for the purpose of analysis, Change and No Change is a key position I compare my data on, regardless of direction or degree of change. This is seen as necessary as the change varies in different epistemic beliefs in the three questionnaires.

Importantly, I may not be able to see the difference in Category Change/Degree Change as something that has enough theoretical background to argue meaningfully in the context of my research. The study of epistemic beliefs is still developing with many grey areas still requiring work. At one stage, I do analyze the data into categories of No Change, Degree Change and Category Change for

fine grained analysis. Besides this, I further classify responses into Low Scores and High Scores to study change from naive to sophisticated positions. These analyses reveal important differences and are useful to draw conclusions, howsoever tentative.

Data is examined as combined i.e. data from all 5 Studies is combined for meta-analysis as well as in separate study groups as the research was conducted.

4.3.3 Statistical Tests Conducted in the Analysis of the BM Data

To study the shift or change in respondents' positions in the combined data, analysis begins with cross tables on the basis of observed scores of individual respondents. This simple analysis reveals a broad picture of change. A Chi Square test is used to see the reliability of change from pre-test to post-test. Furthermore, a non-parametric Wilcoxen rank sum test is used in each city to compare the difference in pre and post stance.

For a more fine grained analysis, whole data is categorized into three categories i.e. No Change, Degree Change, and Category Change. The Mann Whitney test is used in each category of data to compare the changes in both Control and Experimental Groups. Further, another approach is used by comparing the pre and post responses with respect to a low and high stance. In this Section of the study, scores of 1, 2 and 3 are low score and 4, 5 and 6 are a high score. A Chi Square test is used to see the difference in pre and post responses with respect to low and high scores from pre-test to post-test. This is described in the following pages.

Table 4.1: Categories and Corresponding Holistic Scores of Epistemic Beliefs in the CBKH

Categories of Respondents	Holistic Score
Naive Objectivist	1
Critical Objectivist	2
Naive Subjectivist	3
Critical Subjectivist	4
Naive Criterialist	5
Critical Criterialist	6
Sophisticated Integrator	7

The Analysis is described in the following pages with Tables and commentary. This analysis includes the following:

- A frequency distribution of overall combined data to compare change in Control and Experimental Groups.
- Cross tabulation with Chi Square of Low and High scores in overall combined data.
- Odds ratio for Change and No Change stance
- Paired t test in overall combined data for Control and Experimental Groups.
- Wilcoxen Rank Sum Test, to study responses in 5 City Groups.
- Fine Grain analysis of No change, Degree Change, and Category Change.
- Man Whitney Test of No Change, Degree Change and Category Change.
- Table of comparison of difference as significant in each study of 5 City Groups.
- Analysis of Low and High scores in each City Group.

4.3.4 Test of Reliability

Reliability assessment is informed by an Inter-Rater assessment exercise followed by a systematic Internal Evaluation with a concern for rigour. The development of the CBKH category scale was informed by a rigorous process of blind marking, repeated reading and assessment to improve the scale thus making the scores awarded as reliable as possible. The quality of the assessment using the CBKH category scale can be analysed through an examination of the data by interested readers (Appendix-A7). Inter Rater marking and Internal Evaluations are described in detail in separate Sections of this Chapter.

4.3.5 Results of Final Internal Evaluation of Qualitative Assessment and Scoring of Responses to Boscolo & Mason Questions Using the CBKH Category Scale (Appendix-A7)

Match: 10/17 = 58.82
Difference of Degree within Category: 4/17 = 23.52
Category Difference: 3/17 = 17.64
First Internal Evaluation Score: 64.69%
Second Internal Evaluation Score: 76.46%
Third Internal Evaluation Score: 82.00%
Inter Rater assessment and results are described in Chapter 3.

4.4 Statistical Analysis of Response of Participants at Levels of Epistemic Beliefs in Control and Experimental Groups

4.4.1 Comparison of Pre and Post Responses Using Chi Square Test of Independence

Table 4.2 displays the frequency distribution of respondents at a stance. This Table helps observe change in stance at pre-test and post-test in Control and

Experimental Groups. A key difference in the move from a Naive Subjectivist stance to a Critical Subjectivist stance in both sets of data is highlighted.

The Chi Square test for overall BM items of the Control and Experimental Groups reveals that there is a significant difference in stance of respondents in pre and post workshop. The value of Chi Square for the Control Group is 23.73 and for the Experimental Group it is 35.51, and their associated p-value is 0.000 which is less than 0.05. Therefore, it can be assumed from this analysis that Intervention 2 played a role in shifting respondents' stance in the Experimental Group as compared to the Control Group. The difference is seen to be significant in both sets of data. However, the larger Chi Square test statistic for the Experimental Group shows that there is more difference in the Experimental Group as compared to the Control Group.

From Table 4.2, one can see that this difference occurs due to a major shift in the Naive Subjectivist stance in both Control and Experimental Groups. It is interesting to see that before the workshop, 21 respondents which constitute 52.5% of the overall Control Group, appear to be at a Naive Subjectivist stance whereas after the intervention, that is at post workshop, 4 respondents, which constitute 8.9% of the overall Control Group, still remain in this category; while in the Experimental Group there is a 100 % shift in the respondents' stance from the Naive Subjectivist position.

Another difference to be noted is change in the Critical Objectivist stance. The Control Group has 12.5% respondents at pre-test, which nearly doubles to 22.5% at post-test. In the Experimental Group, 16.3% at pre-test, which reduces to 14% at post-test. Respondents in the Control Group nearly double in the Critical Objectivist stance whereas in the Experimental Group numbers decrease in the Critical Objectivist stance. In terms of nature of change this is an important difference.

The difference in change from the Naive Subjectivist to the Critical Subjectivist stance is interesting. In the Control Groups, there is an increase of 4 respondents in the Critical Subjectivist stance whereas in the Experimental Group there is a large increase of almost double in the Critical Subjectivist stance.

In terms of nature of change it appears that most participants move towards sophistication in the Experimental Group. The change is significant in both groups with more significant change being seen in the Experimental Group.

Table 4.2: Frequency Distributions of Participants at Levels of Epistemic Beliefs in Control and Experimental Groups, Pre-Test and Post-Test

	CONTROL			EXPERIMENTAL		
	Pre	**Post**	**Total**	**Pre**	**Post**	**Total**
Naive Objectivist	3	1	4	3	2	5
%	7.5	2.5	5.0	7.0	4.7	5.8
Critical Objectivist	5	9	14	7	6	13
%	12.5	22.5	17.5	16.3	14.0	15.1
Naive Subjectivist	21	4	25	21	0	21
%	52.5	10.0	31.3	48.8	0	24.4
Critical Subjectivist	10	14	24	11	23	34
%	25.0	35.0	30.0	25.6	53.5	39.5
Naive Criterialist	1	10	11	0	9	9
%	2.5	25.0	13.8	0	20.9	10.5
Critical Criterialist	0	2	2	1	3	4
%	0	5.0	2.5	2.3	7.0	4.7
Total	40	40	80	43	43	86

Chi-Square = 23.733
P-Value = 0.000
N = 40

Chi-Square = 35.51
P-Value = 0.000
N = 40

The difference of Chi-Square value in Control - Experimental Groups is 7.78% or 8%. Both Control – Experimental Groups show significant change. P value is 0.00 but the Chi Square is 36.51 in the Experimental Group, or 23.77 in

the Control Group. The difference in nature of change is interesting. More respondents moved up to a higher degree of sophistication in the Experimental Group. A 100% left the Naive Objectivist stance and 14% left the Critical Subjectivist stance. These respondents moved up to Critical Subjectivist stance where a 109% increase is seen. The increase in numbers in Naive Criterialist, Critical Criterialist stance is the same in both Groups.

Figure 4.1: Comparison of Pre and Post Responses in Control Group

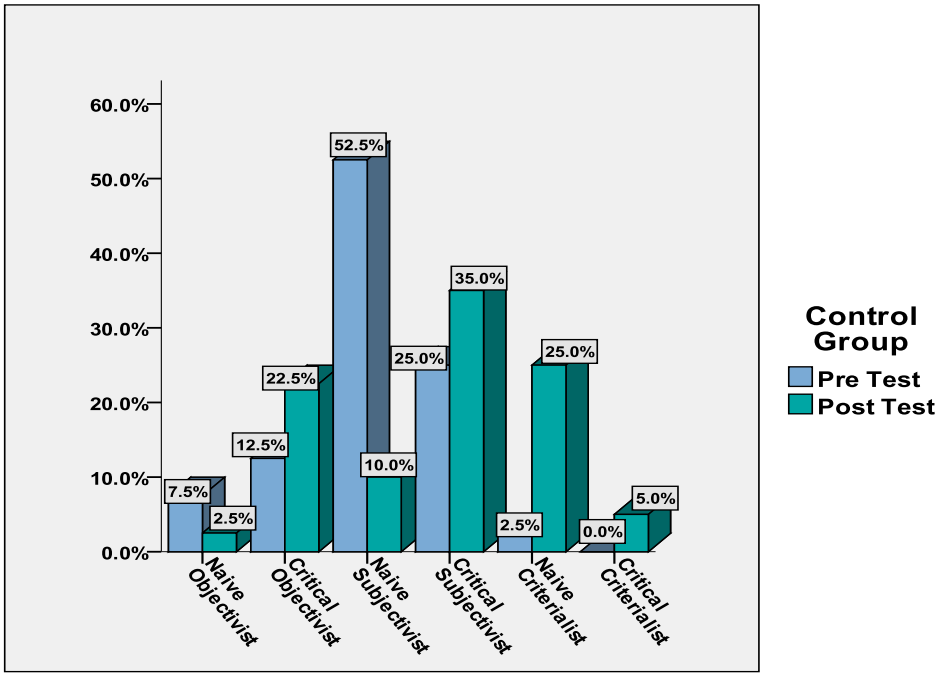


Figure 4.2: Comparison of Pre and Post Responses in Experimental Group

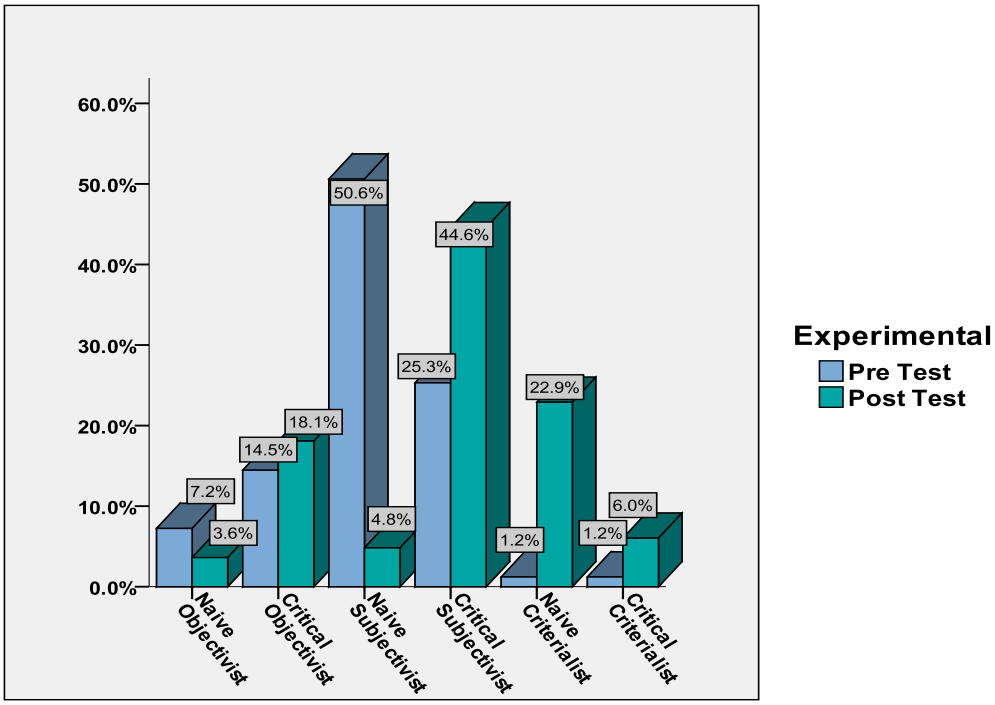
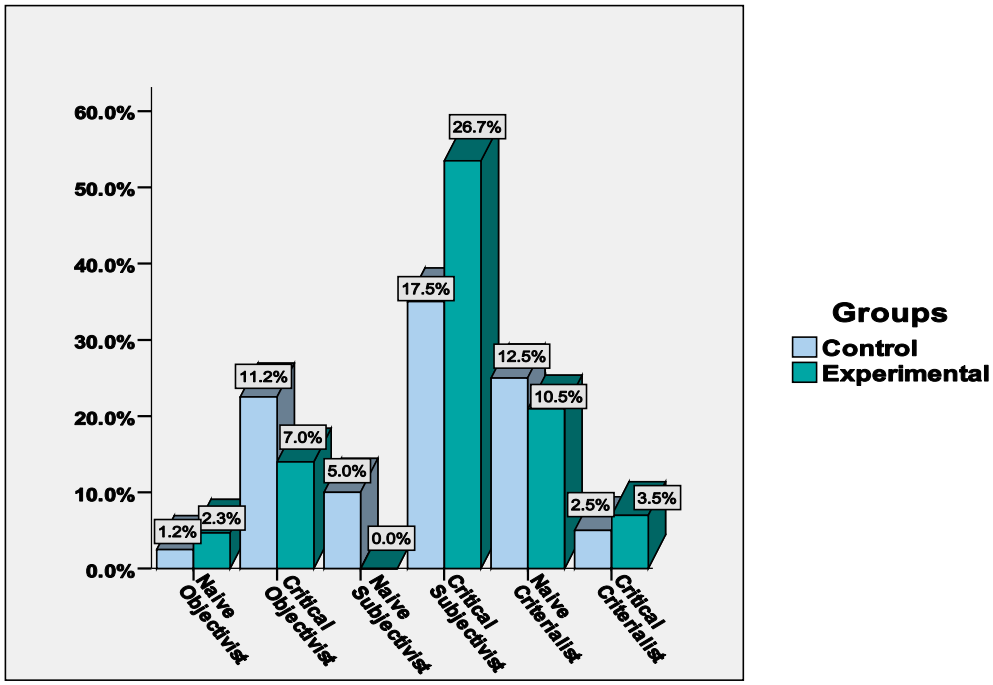


Figure 4.3: Comparison of Post Responses of Control and Experimental Group



4.4.2 Odds Ratio for Change and No Change in Epistemic Beliefs

Next, I categorized the overall combined data into a Change and No Change category. I consider all respondents, who change their stance from pre-test to post-test, fall in the change category. For example, if a respondent is a Naive Objectivist at pre-test and if he/she has any other stance except Naive Objectivist at post-test, he/she falls in the change category. In No Change category, I consider all those respondents, who did not change their stance at post-test and remain at their original position, fall in the No Change category. Now, the whole data is converted into a binary category i.e. Change or No Change. Table 4.3 shows the number of respondents who fall in Change and No Change category in Control and Experimental Groups. In the Experimental Group, 34 respondents out of 43, changed their stance while only 9 respondents did not change their stance. In order to see the likelihood of change than No Change in the Experimental Group, an Odds ratio is calculated for this binary data. The Odds ratio is found to be 1.259 which means there is 1.259, times Odds of change in the Experimental Group than No Change. In other words, I can say that each respondent in the Experimental Group is 1.259 times more likely to be changed than Not Changed. The value of Odds ratio is greater than 1, which indicates that a respondent in the Experimental Group is more likely to change than in the Control Group. The Odds ratio lie within 95% Confidence Interval i.e. 0.452 to 3.512, which shows that the value is not significant. If, however, a 90% or 80% Confidence Interval is considered in this educational context, then the change is significant or at least meaningful.

Table 4.3: Change and No Change Responses in Control and Experimental Group

	Change	No Change	Total
Experimental	34	9	43
Control	30	10	40
Total	64	19	83
	Value	**95% Confidence Interval**	
Odds Ratio	1.259	0.452	3.512

4.4.3 Analysis of Pre and Post Difference in Control and Experimental Group Using Paired t Test

The difference from pre-test to post-tests in both Control & Experimental Groups is significant. In this Section of analysis, a paired t test is used to find the difference in pre and post stance in both sets of data. The mean for the pre-test score in Control Group is slightly different as compared to the post-test score (mean pre-test score= 3.03, mean post-test score= 3.73). The test statistics and significance value (t = - 4.149, degree of freedom = 39, p-value 0.000) show that the difference from pre-test to post-test is significant in the Control Group (see Table 1-B to Table 5-B in Appendix-B). In the Experimental Group, more change (difference) can be seen as compared to the Control Group. The mean of pre-test scores is 3.02 and the mean of post-test scores is 3.93. This difference in mean can be taken to suggest that more change has taken place in the Experimental Group rather than the Control Group. Also, the t statistics and significance value (t = - 5.263, degree of freedom = 42, p-value = 0.000) show that the difference is significant (see Table 1-B to Table 5-B in Appendix -B).

4.4.4 Analysis of Pre-Test and Post-Test Difference in Five Studies for the Control and Experimental Groups

The research is conducted in 5 separate studies in 3 cities. These groups are code named Rose, Tulip, Jasmine, Bougainvillea and Pansy. Each group is analyzed separately to examine the change or difference in responses in individual city groups, then a statistical test of difference is carried out. The data in each group is small, therefore a test of normality is conducted to see the distribution of data. The Shapiro Wilk test of normality suggests that data does not follow normal distribution. Therefore, instead of using paired t-test, a non-parametric test, Wilcoxen Rank Sum Test, is used to see the significance of difference from pre-test to post-test in both sets of data. The test statistics and their corresponding significance value in each city are given in Table 6-B in Appendix-B. The significance value of the test shows that only Rose Group in the Control Group has a significant difference from pre-test to post-test and the remaining four studies do not show significant difference in pre-test to post-test responses. On the other hand, Experimental Groups in four studies, Rose, Tulip, Jasmine and Bougainvillea, show significant difference in their respondents' stance. The result obtained from this analysis is compatible to analysis of combined data that more difference or change is observed in the Experimental Group than in the Control Group.

4.4.5 Estimating Intervention Effect in the Experimental Group in Meta-Analysis of BM Data

The analysis of treatment effect in the BM data is explained in Figure 4.4. The dependent variable 'Y' is response of the respondents in both the Control and Experimental Groups. In Control at pre-test, I observe the Control Group value Y=A, and at the post-test, the Control Group value is Y=E. Similarly, before the intervention in the Experimental Group, the value is Y=B and, after the intervention, Experimental Group value is Y=C. In order to estimate the treatment effect using four pieces of information contained in the points A, B, C and E, from Figure 4.4

the treatment effect is δ = $\overline{CD}$ which is the difference between Control and Experimental Groups at post-test. The estimation of treatment effect is calculated by taking the difference of average values of four points i.e. A, B, C and E as follows:

Treatment effect = δ = $(\hat{C} - \hat{E}) - (\hat{B} - \hat{A}) = (Y_{Experimental,Post} - Y_{Control,Post}) - (Y_{Experimental,Pre} - Y_{Control,Pre})$

Where

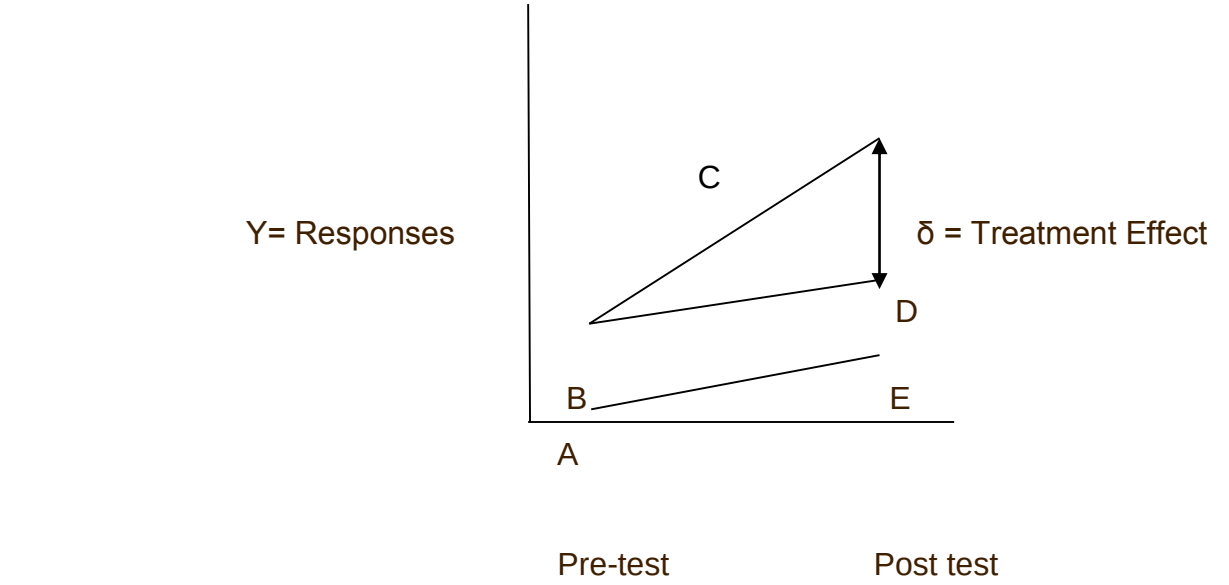
$Y_{Control,Pre} = \hat{A}$ = Sample mean of Y for Control Group at pre-test

$Y_{Control,Post} = \hat{E}$ = Sample mean of Y for Control Group at post-test

$Y_{Experimental,Pre} = \hat{B}$ = Sample mean of Y for Experimental Group at pre-test

$Y_{Experimental,Post} = \hat{C}$ = Sample mean of Y for Experimental Group at post-test

Figure 4.4: Difference-in-Difference Estimation of Control and Experimental Group



From the BM data, the sample mean of four points is found to be as follows:

$\hat{A}$ = 3.025, $\hat{E}$ = 3.725, $\hat{B}$ = 3.02 and $\hat{C}$ = 4.0; so the treatment effect or difference of the difference estimator of the Control and the Experimental Groups is:

δ = (4.0 - 3.725) – (3.02-3.02)

δ = 0.275 – 0

δ = 0.275

So, I quantify the intervention effect in the Experimental Group and it is estimated to be 0.275. This value indicates that the intervention in the Experimental Group does work. If the Experimental and the Control Groups have the same outcome, then the estimated value of difference in difference estimator should be 0 but in the BM data it has a positive value. This positive value somewhat supports the conclusion that the Intervention has a positive effect in the Experimental Group.

The Intervention effect in the Experimental Group in the BM data is also calculated by adding some indicator variables in a simple regression model.

$\boldsymbol{Y_i = \beta_1 + \beta_2}$ **Experimental +** $\boldsymbol{\beta_3}$**PostTest + δ (Experimental × PostTest) +**$\boldsymbol{e_i}$

Where Experimental is the dummy variable, consider 1 when Respondents are in the Experimental Group, and consider 0 when respondents are in the Control Group. In post-test also, a dummy variable considers 1 when respondents are at post-test, and considers 0 when respondents are at pre- test. (Experimental × Post Test) is an intervention variable that measures the intervention effect when respondents are at post-test in the Experimental Group.

Now the regression function is given as:

$\boldsymbol{Y_i = 3.02 - 0.001744}$ **Experimental + 0.7 PostTest + 0.206(Experimental × PostTest) +**$\boldsymbol{e_i}$

S.E	[0.174]	[0.242]	[0.246]	[0.342]
P-value	(0.000)	(0.9943)	(0.0051)	(0.5463)

So, if the respondents are in the Control Group at pre-test, then Experimental = 0 and Post-test = 0 and the model becomes:

Y= $\boldsymbol{\beta_1}$ =3.02 = A (from Figure 4.4); this value indicates that initially the respondents, on average, appear to be Subjectivists in the Control Group at pre-test.

If the respondents are in the Experimental Group at pre-test, then Experimental= 1 and post-test = 0, and Experimental × post-test = 0; now the above model becomes Y= β_1 +β_2 = $3.02 + 0.001744$ = 3.08 = B (from Figure 4.4). This value indicates that the respondents, on average, are at the same position in the Experimental Group as in the Control Group i.e. Naive Subjectivists.

If the respondents in BM data are in the Control Group at post-test, then Experimental = 0, post-test = 1 and Experimental × post-test = 0 and the estimated value of the respondents at this level is:

Y= β_1 + β_3= 3.02 + 0.7 = 3.72 ≅ 4 = E (from Figure 4.4)

This value indicates that respondents, on average, move from Naive Subjectivists to Critical Subjectivists in the Control Group at post-test. Similarly,

when respondents are at post-test in the Experimental Group then their estimated value is calculated as:

Y= $\beta_1 + \beta_2 + \beta_3 + \delta$= 3.02 -0.001744 + 0.7 + 0.206 = 3.94 $\cong$ 4 = C (from Figure 4.4)

This value is very close to 4 whereas in the Control Group the value is 3.72, which is smaller than 3.94. The difference is considered to be very small in a mathematical sense but theoretically, this is an observable value. This value also shows that, on average, respondents in the Experimental Group at post-test appear to be Critical Subjectivists while at pre-test they appear to be Naive Subjectivists.

The suggestion of a difference leads to a requirement of analysis at the level of small studies. This is conducted below. To sum this up, however, a statement is made. The analysis of Low and High stance shows that there is a small positive effect of the intervention in the Experimental Group and the respondents, on average, move from being Naive Subjectivists to Critical Subjectivists.

4.5 Fine Grain Analysis of Change in Respondents Stance

4.5.1 Fine Grain Analysis of Classes, No Change, Degree and Category Change

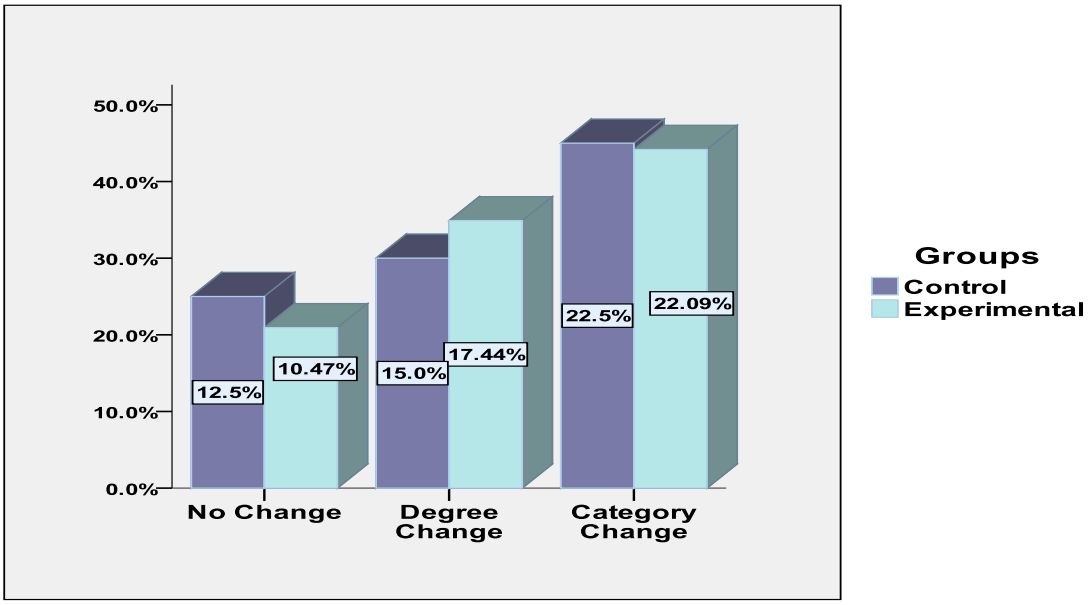
Change can be ordered as a transition within a category i.e. within a stance, say of an Objectivist, the respondent can be Naive or Critical on the basis of a beginning of doubt. The respondent can move forward or in reverse within the category. This is labeled Degree Change. Category change is awarded as a move from a category such as an Objectivist stance to a Subjectivist stance or a Criterialist stance, or in reverse in any order. Table 4.4 shows the changes in stance with respect to three classes No Change, Degree Change, Category Change. I categorized the BM score into the three classes, No Change, Degree Change, and Category Change. I considered those respondents, who did not change their position at post-test, to fall in the class of No Change. For example,

if a respondent appeared to be a Naive Objectivist at pre-test, and at post-test he/she again appears to be Naive Objectivist, I consider that he/she falls into the No Change class.

Table 4.4: Analysis No Change, Degree Change, Category Change in Control and Experimental Groups

Category Scale	No Change		Degree Change		Category Change	
	Control	Experi-mental	Control	Experi-mental	Control	Experi-mental
Naive Objectivist	1	0	2	2	0	1
Critical Objectivist	2	1	0	0	3	6
Naive Subjectivist	3	0	10	0	8	0
Critical Subjectivist	3	0	0	13	7	8
Naive Criterialist	1	7	0	0	0	4
Critical Criterialist	0	1	0	0	0	0
Total	10	9	12	15	18	19

Figure 4.5: Comparison of No Change, Degree Change and Category Change Stance of Control and Experimental Groups



These respondents display no effect of the intervention at all. This can also be discussed as a possible resistance to change.

In Degree Change, those who shift their stance within *a position* of Objectivist, Subjectivist, Criterialist from a firm position on the stance (I call this a naive position on the CBKH) to a beginning of doubt (I call this a critical position) are said to have changed a degree. These transition stages are described in detail in Chapter 3 on the CBKH Categories. Change from 1 to 2 i.e. from Naive Objectivist to Critical Objectivist or from 3 to 4 i.e. Naive Subjectivist to Critical Subjectivist falls in the Degree Change class.

For the Category Change class, I considered those respondents who shift their position from Objectivist to Subjectivist or Criterialist in any order or direction. This can be a move towards greater sophistication or towards a more naive stance. For example, if a respondent appears a Naive Objectivist De Vaus (2001) in pre-test and at post-test he/she shifts to a Naive Subjectivist (Score 3), I consider that he/she has made a Category Change. Another example is, if a person is at a Naive Subjectivist category and goes back to a Naive Objectivist category that is

also a Category Change. This appears complicated but it is necessary as the change itself is complex and I have tried to observe it as systematically and as much in depth as possible.

It is crucial to remember that for the purpose of my research, Change and No Change are the deciding difference between success and failure of the Intervention to effect the epistemic beliefs of respondents. This difference is used to answer the research question. Having said that, change is examined in some depth to add to my understanding of the nature of change and the process of change. The data is very interesting and leads to many avenues of study.

Table 4.4 displays a comparison of No Change in stance, Degree Change, and Category Change, between Control and Experimental Groups. The slight difference in favor of the Experimental Group is consistent with other analysis. There is also a tiny difference in No Change in the Experimental Group with less people resisting change in the Experimental Group.

4.5.2 Results of Mann Whitney Test of Classes of Change, No Change, Degree Change, and Category Change

The results obtained from Fine Grain Analysis in No Change, Degree Change and Category Change in each category scale i.e. Naive Objectivist to Critical Criterialist are then compared in the Control and Experimental Groups.

Again, the sample drawn from this analysis is very small, therefore instead of using independent t test a non-parametric Mann Whitney sum test is used to compare the changes in Control and Experimental Groups.

The result or test statistics obtained from this test indicates that in No Change category, the change in both sets of data is reliable and significant. The mean rank of the Control Group is 6.65 while the mean rank of the Experimental Group is 13.72, and shows more change in the Experimental Group (see Table 7-B in Appendix B). The significance value of this test is 0.004, which shows that there is significant difference in both groups in the category of No Change (see Table 8-B in Appendix B).

For the Degree Change category, the result is also similar to the No Change category. The significance value of the Mann Whitney test indicates that the difference in both groups is reliable. Also, the mean rank of the Control Group is 8.33 and the mean rank of the Experimental Group is 18.53, which shows a large difference in the mean of Degree Change for both groups (see Tables 9-B and 10-B in Appendix B). The larger mean rank of Experimental Group indicates that most respondents in Experimental Group fall in the Degree Change category.

For Category Change the difference in Control and Experimental Groups is not reliable or significant. The mean rank for both groups is almost similar to each other. In the Control Group, the mean rank is 17.56 and for the Experimental Group the mean rank is 20.37. The significance value (p-value) for Category Change data is 0.407, which shows no significant difference in both groups for Category Change (see Table 11-B in Appendix B).

4.5.3 Fine Grain Analysis of Low and High Stance for Control and Experimental Groups in Overall Combined Data

In this Section of Fine Grain Analysis, the Epistemic scores on the CBKH ranging from 1-3 are classed as Low scores (from Naive Objectivist to Naive Subjectivist). Scores ranging from 4-6 are classed as High scores (Critical Subjectivist to Critical Criterialist). From Table 4.5, it is possible to recognize that most respondents fall in the Low category at entry pre-test in both sets of data. In the Control Group, 29 respondents constitute 72.5% of overall pre respondents that fall in the Low category while 11 respondents 27.5% fall in the High category. Similarly, in the Experimental Group, 31 respondents i.e. 72.09% of pre respondents appear to have a Low stance. Only 12 respondents in the pre-test Experimental Group show a High stance. This similarity in responses at entry pre-test is important and encouraging as it suggests that both groups were similar at outset even if the selection was non-random or not matched. An examination of independent studies suggests finer differences between groups although the general level and pattern is similar (see Table 12-B in Appendix B).

It is very interesting to see the pattern of cange in post stance of both Control and Experimental Groups. In the post-test, the picture is opposite to the pre-test. In the Control Group, 13 respondents that compose 32.5% of overall post respondents, have a Low stance. 27 respondents, make 67.5% of overall post respondents and fall in the High category. The picture is clearer in the post-test scores of the Experimental Group; 8 respondents in the Experimental Group fall in the Low category whereas 35 respondents i.e. 81.4% of overall post experimental respondents, fall in the High category. This significant change in stance from Low to High stance in the post-test of both sets of data indicates that Interventions 1 and 2 do work in their own ways and are compatible with my hypothesis; that more change could be observed in the Experimental Group than the Control Group. In order to support this statement, I used a Chi-square test to see the reliability of change in the Control and Experimental Group. In Table 4.5, the Chi square value for the Control Group is 12.832, and shows a difference in pre-test and post-test. The significance value for this set of data is 0.000, which is very small and shows difference or change is significant in pre-test and post-test responses. In the Experimental Group, the Chi square value is 24.819, which is almost double the Chi square value of the Control Group. This value indicates that there is a large difference in change in the Experimental Group as compared to the Control Group. Also, the significance value for this set of data is 0.000, which shows that Change is reliable and significant.

Table 4.5: Analysis No Change, Degree Change, Category Change in Control and Experimental Groups

Category	Control Pre test	Control Post test	Experimental Pre test	Experimental Post test
Low	29	13	31	8
%	72.5	32.5	72.1	18.6
High	11	27	12	35
%	27.5	67.5	27.9	81.4
Total	40	40	43	43
	Chi square = 12.832		Chi square = 24.819	
	DF =1		DF =1	
	P-value 0.000		P-value 0.000	

Figure 4.6: Comparison of Low and High Stance Control Group

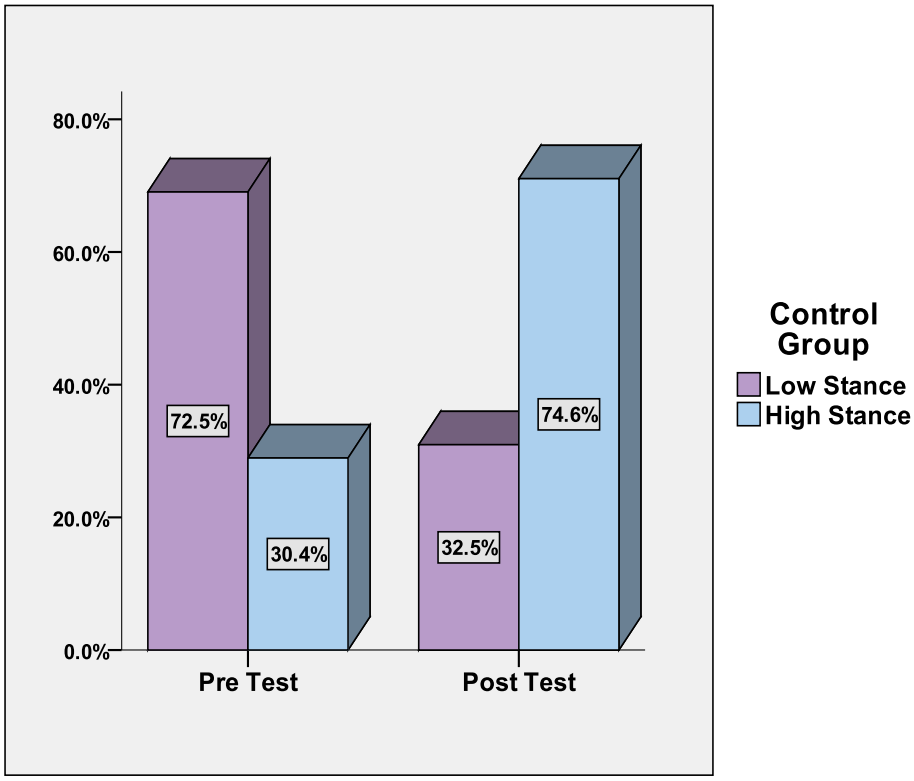
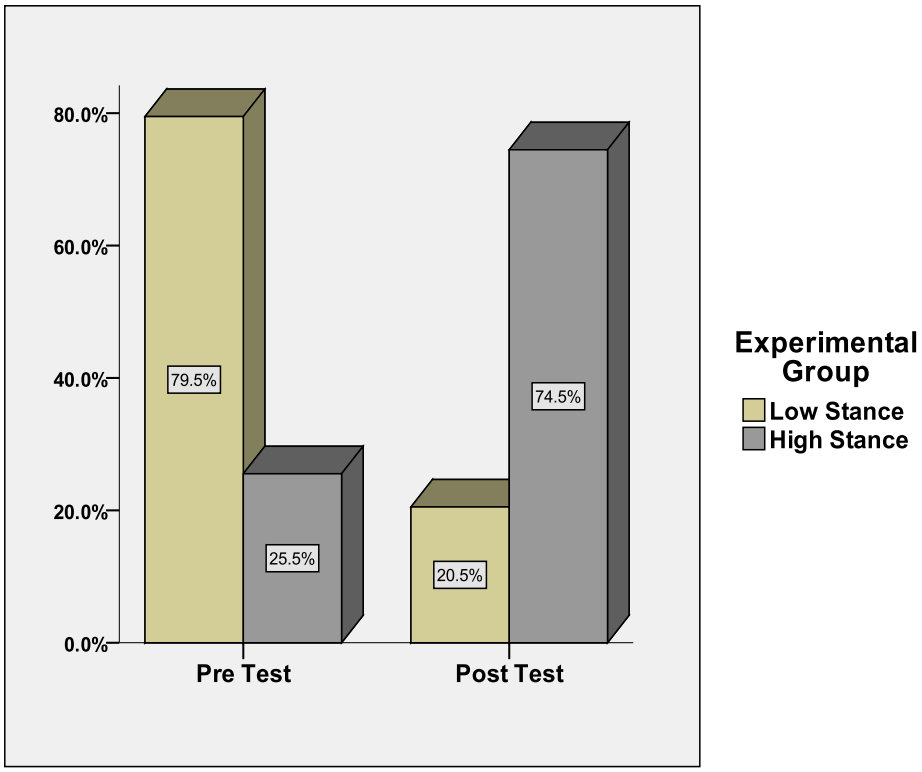


Figure 4.7: Comparison of Low and High Stance Experimental Group



4.5.4 Fine Grain Analysis of Low and High Stance in the Five Study Groups

An analysis of Low and High stance in the Control and Experimental Groups in the 5 study groups was conducted. At pre-test in the Control Group, most participants are at Low positions (1-2-3) on the Category Scale except in Rose, where Lows and Highs are more balanced. At post-test in the Control Group, A, B, and E present a matching picture with most participants moving to the High Group and lower stances being vacated. In the Experimental Group at pre-test, most groups present a picture of Low scores except Jasmine, where 53.8 % show High scores. In the post-test of the Experimental Group, the picture is reversed with all except one showing significantly High scores. Pansy differs in this respect. Pansy has no participant at High scores at pre-test, which could be the reason why in a comparison of Low and High scores, the group is at a disadvantage. It is interesting to note the similarity in the two groups at Total. The percentage of Low scores at pre-test is 72.5 in the Control Groups and 72 in Experimental Groups. There are

27.5 in the High scores at pre-test of the Control Group and 27.9 in the Experimental Group. The difference in total change to High scores is very different at post-test between Control and Experimental Groups. 81.4 % are at High scores in the Experimental Group as compared to 65% in the Control Group. A change can easily be seen in terms of observed counts of Low and High category in both sets of data. This result is studied further with an estimate of Effect.

Table 4.6: Analysis of Low and High Stance in each City's Study Group for Control and Experimental Groups

Study Groups	Control				Experimental			
	Pre		Post		Pre		Post	
	Low	High	Low	High	Low	High	Low	High
Rose	6	5	1	10	5	1	1	5
%age	54.54	45.45	9.09	90.9	83.3	16.6	16.6	83.3
Tulip	5	0	1	4	4	2	1	5
%age	100	0	20	80	66.6	33.3	16.6	83.3
Jasmine	9	2	7	4	6	7	1	12
%age	81.8	18.1	63.6	36.3	46.1	53.8	7.7	92.3
Bougainvillea	6	2	4	4	9	2	1	10
%age	75.0	25.0	50.0	50.0	81.8	18.2	9.1	90.9
Pansy	3	2	1	4	7	0	4	3
%age	60.0	40.0	20	80	100	0	57.1	42.9
Total	29	11	14	26	31	12	8	35
%age	72.5	27.5	35	65	72.0	27.9	18.6	81.4

The data of the five study groups was analyzed to identify more change towards High scores in each. There is a pattern of more Low scores at pre-test while there are some differences in proportion between groups. However, the proportion of difference at post-test is not always concurrent. Groups differ in effects. The analysis above compares a sum of moves towards High scores in

each group. Table 4.6 shows that there was more change in the Experimental Groups overall than in the Control Groups. This is consistent with the higher change in the Experimental Group in the combined data and the higher %age of High scores in the Experimental Group at post-test.

4.5.5 Difference-In-Difference Estimator after Excluding Pansy

Table 4.7: Difference-In-Difference (D.I.D) Estimator - City Wise

City	D.I.D
Rose	0.590909
Tulip	0.1
Jasmine	0.454545
Bougainvillea	0.443182
Pansy	-0.3619

To measure the intervention effect in each group, a Difference-In-Difference Estimator is calculated by the following formula:

Treatment effect = δ = $(\hat{C} - \hat{E}) - (\hat{B} - \hat{A}) = (Y_{Experimental,Post} - Y_{Control,Post}) - (Y_{Experimental,Pre} - Y_{Control,Pre})$

Treatment effects of all groups except Pansy show that there is a positive effect of the Intervention in Experimental Groups. 4 out of 5 Experimental Groups respond better as compared to the Control Groups. Only Pansy shows that the Intervention did not work and its Control Group performs better than the Experimental Group.

The Pansy group appeared to be an outlier compared to the other groups. Therefore, to explore this, I excluded Pansy in the analysis and calculated a Difference- In-Difference (D.I.D) for the remaining four groups. D.I.D Estimator for the four groups is found to be 0.35, which is greater than the D.I.D value of the overall five study groups. Also, I ran a Difference-In-Difference regression model for the four groups as I have carried out in the meta-analysis of all Cities. The significance value of intervention effect in this model is found to be 0.000, which is better than the previous model. This is an important comparison and provides support to the hypothesis that there is a relationship of a perception of agency and change in concepts.

4.6 Summary of Analysis of Data from Responses to BM (Boscolo and Mason) Questions

Notes:

a) Responses were analysed using the CBKH rubric. Qualitative descriptors were awarded numerical scores on a scale of 1-7. Scores ranged between1-6, no individual was found at 7.

b) Analysis is of overall combined data of 5 studies and separately of 5 independent studies.

4.6.1 Reliability of the Instrument

- First internal Evaluation score = 64.69%, and percentage of mismatch is 35%
- Second Internal Evaluation Score= 53% and percentage of mismatch is 47%
- Third internal Evaluation score = 82% and percentage of mismatch is 17%

4.6.2 Frequency Distribution Table with Chi-Square Test

The Control Group Chi-Square test statistic of change from pre-test to post-test is 23.733 and p-value is 0.000. Experimental Group Chi-Square test statistics is 35.51 and p-Value is 0.000. Change is significant for both groups but the Chi-Square statistic is larger for the Experimental Group. The difference of values between Control and Experimental Groups is, however, not significant.

4.6.3 Odds Ratio for Change and No Change

Odds ratio is 1.259. There is more likelihood of change in the Experimental Group than the Control Group.

4.6.4 Pre-Test and Post-Test Difference in Both Groups Using a Paired t Test

The difference is significant from pre-test to post-test in both groups. Control Group test statistic and significance value is (t = - 4.149, degree of freedom = 39, p-value 0.000). In the Experimental Group, change is also significant. (t = - 5.263, degree of freedom = 42, p-value = 0.000) However, more change has taken place in the Experimental Group.

4.6.5 Pre-Test and Post-Test Difference in 5 Studies Using Wilcoxen Rank Sum Test

Test results indicate that only 1 out of 5 groups in the Control Groups has significant difference in responses while in the Experimental Group, 4 groups out of 5 show significant difference from pre-test to post-test.

4.6.6 Estimating Intervention Effect Using Difference-In-Difference Estimator

Effect of the treatment is 0.275. This positive value indicates that Intervention in the Experimental Group did work to make a difference in change.

4.6.7 Fine Grain Analysis of Change in Stance (No Change, Degree Change and Category Change) with a Mann Whitney Test

The difference is significant between Control and Experimental Groups in the No Change and Degree Change categories while not significant in the Category Change category.

4.6.8 Analysis of Low and High Stance in Overall Control and Experimental Groups

A Chi square test of independence was used to see the significance of change (difference) in Low and High stance. The Chi square test statistics in Table 4.2.6 shows significant difference in Low and High stance in the Control Group. While more significant difference is found in terms of Low and High stance in the Experimental Group.

4.6.9 Fine Grain Analysis of Low and High Stance in 5 Independent Studies Using a D.I.D Estimator

D.I.D estimates are found to be positive in the first four groups while negative in the last group. Intervention 2 has a positive effect in 4 groups out of 5.

4.7 Analysis of Data from the Likert Scale, the Belief about Learning Teacher History Questionnaire (BLTHQ)

4.7.1 Questions for Analysis of the BLTHQ Data

- What is the stance of participants in terms of agreement or disagreement with items reflecting the Categories, Copier, Borrower, Criterialist, at pre-test?
- What is the stance of participants after Intervention at post-test?
- Do participants change their stance from pre-test to post-test in both groups, Control and Experimental?
- What is the degree of change?
- What is the nature of the change and the direction of the change?
- Is there a difference in numbers at change from pre-test to post-test between Control and Experimental Groups?
- Is there a difference in degree of change or nature of change between Control and Experimental Groups?

4.7.2 Reliability of the Instrument

Reliability (internal consistency) of the BLTHQ items in pre and post -test for both Control and Experimental Groups is measured by using Cronbach's alpha. The Cronbach alpha for the Control Group at pre-test is 0.562 indicating very weak consistency, and at post-test, the reliability is increased to 0.752 showing fairly good reliability of the items. In the Experimental Group, the reliability of the instrument at pre-test is found to be 0.612 and, at post-test, the reliability is found to be 0.639, which can be taken to mean that the reliability value is in an acceptable range in the Experimental Group. It must be kept in mind that there are only 21 items in the Likert Scale and these can be grouped into three different levels of beliefs. Factor Analysis does show more items loading on Factors at pre- test than at post- test.

There is some doubt in my mind regarding the Cronbach alpha for the BLTHQ scale. The items are divided into three different sets according to the three categories of belief, the Copier, Borrower, and Criterialist. The Cronbach alpha is a measure of the internal consistency among the items. Cohen et al. (2007, p.506). Cohen et al. (2007) refer to Bryman and Cramer (1990:71) that the reliability value is acceptable at 0.8 although others suggest that it is acceptable at 0.67 or above. Keeping the latter as a reference the yield quoted above may be considered adequate.

Apart from this test, the Factor Analysis is presented for discussion. Importantly, the other measure of change and epistemic stance, the Boscolo & Mason questions, display similar, consistent results. A number of tests conducted of data from the BLTHQ and presented here also show consistent if not significant results.

4.7.3 Arrangement of Data

4.7.3.1 Change as Observed in Responses to the BLTHQ Items is Expressed in the Following Arrangement:

A binary model of Change and No Change, where all data is arranged according to the difference in scores obtained in pre-test and post-test. If the score on the Likert Scale at pre-test is the same as obtained at post-test, this is seen as No Change. If the score is different at post-test, regardless of where it is on the scale of 1 to 5, it is seen as change. In this arrangement, all change is change as opposed to No Change as resistance to change if a pattern appears to emerge.

A move from agreement on the item to disagreement on the item at post-test or vice versa is seen as a change in category on a single item. Change as change in category, Copier, Borrower, or Criterialist, as agreement or disagreement with a number of items in the same category, will be seen as a change in stance. Of the 21 items, 9 are said to be Copier items, 4 are said to be Criterialist items, and 8 are said to be Borrower items. If a respondent is seen to

be agreeing with most Copier items and disagreeing with Borrower or Criterialist items, that respondent may be said to be holding a Copier stance. This position could change. A change in category selected at pre-test by the respondent to another at post-test will be seen as a change in category. Change can be change in degree of stance e.g. from a score of 1 as strongly agree to a score of 2 as agree with the item that denotes the stance.

4.7.3.2 Results and Discussion of Analysis on Cross Tables and Chi Square

In order to observe and understand the pattern of change or difference in pre and post responses, Cross tables of all Items were made and studied in depth. The Chi Square test statistics in most items are found to be greater in the Experimental Group as compared to the Control Group. This indicates that more change (difference) took place in the Experimental Group. The significance value indicates that only 3 items out of 21 items in the Experimental Group are significantly different in pre-test to post-test responses. The change, however, appears to be visible on the cross table. All cells are not homogenous from pre-test to post-test in both groups and movement can be observed in degree of stance if not easily from categories. This needed to be quantified in some way and, if there was a pattern to this change, it needed to be discovered. I, therefore, searched for options of analysis in order to observe and understand the pattern of change or difference in pre-test and post-test responses. Significance value of the Chi Square reveals that change is insignificant in most items of both groups except for 3 items. In these 2 Items of the Criterialist stance, and 1 Item of the Copier stance, more significant change has taken place in the Experimental Group only. As the Criterialist stance is seen to be the key, this difference between Control and Experimental Groups is considered meaningful. An instrument effect could be that as there are only 4 items out of the 21 of the Criterialist stance in the BLTHQ, and large change in both groups is seen in the Criterialist stance mainly, it can be suggested that with more items in this stance, change could be more significant

on the BLTHQ. The cross tables analysis of most items shows that the data does not appear random or confused for the most part. There is a clear pattern that is apparent in the consistency of responses falling in easily discernible bands for most items in both groups. It may be possible to assume that the questionnaire, the BLTHQ, has largely been able to identify the participants' beliefs. The Exploratory Factor analysis has interesting results.

4.7.3.3 Quantifying Difference of the Difference in Change as seen on Cross Tables

It was seen as important to know the difference of the difference in change from pre-test to post-test between the Control and Experimental Groups in order to support an answer to the research questions. To quantify the amount of change in both sets of data, an absolute difference of percentages in each category from pre-test to post-test is taken in both sets of data, Control and Experimental. The reason for taking absolute difference is that I only need the magnitude of change, therefore I have ignored the sign and considered only the absolute value. Then, a difference of these differences is taken.

In Table 4.16, ΔC% represents change from pre-test to post-test in the Control Group while ΔE% represents change from pre-test to post-test in the Experimental Group. The sum of difference is calculated separately in both Control and Experimental Groups and finally, a difference of this difference is taken. The overall change in the Control Group is 20% while in the Experimental Group the change is 32.50%, which shows that more change has taken place in the Experimental Group as compared to the Control Group.

This process is repeated for each item and is stated in the column as in Item one. By studying the pattern of these differences for all items, it is possible to see that out of nine Copier items, six items show more change in the Experimental Group. All four Criterialist items show more change in the Experimental Group at post-test. Out of eight Borrower items, five show more change in the Control Group.

To summarize, Copier and Criterialist items show more change in the Experimental Group, and Borrower items show more change in the Control Group. Out of a total of 21 items, 13 items show more change in the Experimental Group and 8 items show more change in the Control Group. This analysis of the difference of the difference helps to quantify the difference in groups and may support the notion that change took place in both groups, Control and Experimental, while more change took place in the Experimental Group. Why there is a difference in change in Borrower items is hard to understand. Theoretically, the pattern of more change in the Criterialist items towards greater agreement or towards a more Criterialist stance stands to reason but the difference in the Control Group on the Borrower stance alone needs further analysis. It is certainly consistent with theory that a move from a Borrower stance, which is related to a Subjectivist position, is a move towards more sophistication. In that sense, it can be seen as consistent with a move towards a more Criterialist stance in both sets of data.

There is a similar move in the Borrower stance in the Experimental Group but it is not significant. The move in the Experimental Group in the Criterialist stance is important as the Criterialist position is key in the scale of epistemic beliefs and the Intervention introduced activities and discussion related to this. The group changed significantly more on this stance and changed towards greater agreement with the Criterialist stance.

Data indicates an emerging consistency in responses in both groups. There is evidence of change in belief as degree of agreement or disagreement rather than change of category of belief. On the whole, the pattern is of consistency in category from pre-test and post-test. There is more change in degree of agreement or disagreement within the category rather than change from category. There is evidence across the tables of change in degree from a strong to a simple agreement or disagree at a stance. This is different from what the BM results suggest where change from categories is evident across the board. The stance that is apparent is conceptually meaningful when seen in the light of data from the BM questions.

The cross tables display the percentages and frequencies that support this conclusion. An analysis of stance and consistency of stance is conducted. Most respondents disagree with most Copier items in both sets of data, Control and Experimental, and more so, at post-test. Respondents agree with Criterialist items, more so in the Experimental Group at post-test, and to some extent, disagree on Borrower items thus demonstrating a more Criterialist stance. This is common in the main in both groups with differences in change in degrees of agreement and disagreement. More change appears to have taken place in the Experimental Group and this is significant in the Criterialist items, which is also important. The analysis of stance as percentages of agreement and disagreement in the cross tables and of the consistency of stance support the findings.

Where I had expected that change would be evident in both instruments making triangulation possible to support the findings, has been supported. Results also appear to strongly suggest that the history focused questions of the BM did work to unravel history related epistemic beliefs while the BLTHQ did access the learning and teaching beliefs held by teachers. If these results are reliable, and it appears so to some extent, interesting questions are raised in reference to the similarity in change as mapped by the two questionnaires. Are the teachers' epistemic beliefs about knowledge in history and their understanding about pedagogical practices consistent? Was this the case because these were practice related belief embedded in school culture or did the Intervention make a difference?

Interestingly, employing the data from each instrument separately to analyze what changed, and did not change, whether the intervention worked to target those particular beliefs or not, can be seen to be an effective strategy for analysis of the BLTHQ and BM data. The statements of the BLTHQ and BM questions are different as is the nature of the beliefs that each targets, yet each instrument displays change from pre-test to post-test in its own way. This can serve to inform research in the field.

The BM questions target theoretical ideas and the BLTHQ statements are set in pedagogical practices of history teaching. The responses in both instruments

provide useful data regarding these practices and the beliefs that may underpin them. Epistemic beliefs are complex and relate to domains and contexts as well as other factors as research suggests. The stance of participants and change as seen in these two instruments can provide support to descriptions of teacher beliefs and implications for change.

The change, as seen post interventions, has implications for the change in syllabus and assessment as planned by Cambridge International Examinations. Schools and teachers as well may see the emerging story as useful in informing both teacher education requirements as well as the possibilities for managing change.

4.7.4 List of Statistical Analysis of the BLTHQ Items to Identify and Analyze Change

In order to examine the change in pre-test and post-test and the difference in change in both sets of data, Control and Experimental, I have used various statistical techniques. The reason for conducting a range of analysis was to limit possibilities of error and look for finer grain in the analysis.

The BLTHQ is constructed on a scale from Strongly Disagree to Strongly Agree. For this Instrument, I have used 1 as Strongly Disagree, to 5 as Strongly Agree. The data is at ordinal level with no specific assumptions, therefore non-parametric statistical tests are used to analyze the data. These are described below.

4.7.4.1 Cross Table and Chi Square

First, individual items are analyzed by means of cross tabulation and a Chi Square test to see the difference between observed and expected frequencies in each set of data. The Chi Square test statistic in each item is calculated to see the difference in pre-test and post-test responses with respect to each category (Strongly Agree to Strongly Disagree).

One of the basic assumptions of Chi Square is that 20% of expected count in each is not less than 5. If the 20% of expected count is less than 5, then the Chi Square asymptotic significance value can be misleading. In this case, the Fisher Exact significance value is used instead of asymptotic significance value. In my study, several items have expected count less than 5 in over 20 % of cells. So, I report Fisher Exact significance value where it is needed and ignore asymptotic significance value. The Chi Square test is used to see the difference in two sets of questionnaires at different levels of agreement and disagreement. An effect size of this Chi Square value is calculated to report the measure of association in terms of weak, modest, moderate and strong association. Since the dependent variable is ordinal in nature (Strongly Agree to Strongly Disagree) and the independent variable is dichotomous (pre-test and post-test), Cramer's V measure of association is appropriate to calculate the effect size (Appendix-B).

4.7.4.2 Analysis of Stance from Cross Tables with Percentages at Agreement and Disagreement

The Cross Tables display the frequencies of agreement and disagreement on each item. In order to discuss the stance of participants, a preliminary analysis of percentages at agreement and disagreement is carried out. In places in some items, percentages of Agree and Strongly Agree are summed to discuss the stance of participants. By simple addition of the percentage of Agree and Strongly Agree, a sum of agreement is taken. This is repeated for Strongly Disagree and Disagree. Thus, a percentage sum is taken to point towards a stance of larger numbers of participants. A few items are displayed in this file. The remaining can be seen in the Appendix 14-B to 27-B.

4.7.4.3 Analysis of the Difference of the Difference

To carry out a closer analysis of change, I used another mathematical method based on observed responses (frequencies) in the Cross Table. The

process of this method is in two stages. First, I took the absolute difference of pre-test to post-test responses in each category for each group, Control and Experimental. Next, a sum of these absolute differences was taken separately in both Control and Experimental Groups. The difference of this sum from Experimental to Control Groups represents total change in each item.

To explain, the cross table displays the percentage of responses for each category, Strongly Agree to Strongly Disagree. Using Excel, the percentage at pre-test was subtracted from the percentage at post-test for each category in the Control Group. This was repeated in the Experimental Group. Taking absolute numbers and disregarding the negative and positive signs, a sum of this absolute number is taken. This sum of the Control Group was then subtracted from that of the Experimental Group. The remainder was the difference of the difference between pre-test and post-test scores between Control and Experimental Groups. This is displayed in Table 4.16.

This analysis will help me answer research questions on the basis of whole groups, pre-test and post-test and Control and Experimental, but it will not tell me much about how individuals changed. The analysis of Cross Tables also tends to focus on large numbers in cells even if they are technically not 'most'. A pattern horizontally across the Table 4.16 shows where a band of large numbers is formed, often from Control to Experimental Groups. This also misses the numbers scattered at other points of agreement and disagreement, which is a disadvantage of the method. However, it does help to answer the question of where broadly change took place.

4.7.4.4 Wilcoxen Rank Sum Test

A Wilcoxen Rank Sum Test, a non-parametric test, is used in each of the 21 individual items for both sets of data, Control and Experimental, to see the difference in pre-test and post-test responses. The test is also used in the smaller sets of data of individual City groups to study the change in pre-test to post-test responses in Control and Experimental Groups in depth.

4.7.4.5 Odds Ratio

In the next stage of study, the responses of overall respondents are categorized into Change and No Change categories in both Control and Experimental Groups. The data is converted into binary categories i.e. Change or No Change. The binary data is then analyzed by means of Odds Ratio to find out the Odds of change in the Experimental Group rather than No Change or Odds of change in the Experimental Group rather than the Control Group.

4.7.4.6 Analysis of Stance

It was considered important to identify the stance of participants in order to describe the change from pre-test to post-test and relate this to the Interventions provided. The focus of the study is conceptual change, and epistemic belief change is the particular conceptual change of interest in my study. By itself, exploring the epistemic stance of teachers of history is a useful endeavour in the light of the expected change in the Cambridge O' Level History Syllabus and its assessment in Pakistan. An Exploratory Factor Analysis is conducted followed by weighted averages and Consistency scores.

4.7.4.7 Exploratory Factor Analysis

An Exploratory Factor Analysis based on a principal components extraction method is carried out to identify the structure of data as well as to explore the factors that emerge from the data. A similar analysis was conducted by the authors in the US and it is reported by them e.g. Maggioni et al. (2004), I attempted the Factor Analysis in order to see how the instrument worked with my sample and if this could be seen to support my findings from the other analysis in any way.

Factor Analysis was conducted of separate sets of data of pre/post, Control and Experimental Groups and followed by an analysis of the complete data set of 166 response sheets. Factors that emerged remain 7 or 8 initially but those 3 that explained more variance were taken. The analysis of the overall data provided a

more coherent picture as 3 Factors of a single stance did emerge with an exception of 1 item in each.

An unexpected outcome of carrying out the Factor Analysis turned to be an observation of the difference in the Factors that emerged at pre-test and post-test taken in the Control and Experimental Groups. Where Factors of a single stance did not emerge at pre-test, at post-test, data of the Experimental Group showed a change with a Factor of a single criterialist stance emerging. This is described in the Section 4.7.12 on Factor Analysis.

4.7.4.8 Weighted Score and Consistency Score of Epistemic Beliefs

A final analysis of responses to the BLTHQ is conducted to identify the epistemic stance of respondents and describe the change within this. Weighted scores and Consistency scores are calculated for each respondent. As there are a total of 83 participants, 40 in the Control Group and 43 in the Experimental Group, an overview is required; therefore, a Table of scores is presented in Table 4.26. A discussion is provided below the Table 4.26. The Weighted score is calculated to assess the magnitude of change from pre-test to post-test. The Consistency score is calculated to assess the consistency of respondents on a stance. To explain, a respondent, to be a Criterialist, should in principle agree with criterialist items and disagree with Copier and Borrower items. The authors set a high bar, expecting 90% before a respondent can be assigned to the stance. This is seen in the data as over 90% in actual terms agree with the items of the criterialist stance; however, because they also either agree with the Copier stance or the Borrower stance, consistency is denied. That is an interpretation however, as it is hard to say *why* respondents did so. More qualitative data in a follow up study could help to explain this phenomenon.

4.7.5 Description of Analysis of Change

4.7.5.1 Cross Table of BLTHQ Items

The stances of respondents are categorized into the Copier, Borrower and Criterialist Stance as described by the authors, Maggioni et al. (2004) BLTHQ items reflect a Criterialist (4,11,12,13), Borrower (14,15,16,17,18,19, 20, 21) and Copier (1,2,3,5,6,7,8,9) stance. Cross tables of all 21 BLTHQ items are constructed individually. The percentage difference, Chi Square statistics value, its significance value, effect size and co-relation among pre-test and post-test responses of both Control and Experimental Groups are calculated for each item. In this Section, only items 1, 2, 7 of the Copier, items 4,11,12,13 of the Criterialist and item 16 of the Borrower stance are presented and discussed in detail. The remaining items are available in Appendix-B (see Tables 14-B to 27-B).

These items are selected for discussion for the following reasons. The first item shows responses scattered in cells and is a good example of such items. Item 2 is one of those that show significant change and that happens to be in the Experimental Group. The Criterialist items 4, 11, 12, 13, are selected because the Criterialist stance is important being the point of reference for a more sophisticated stance and where most change has taken place. Significant change has taken place in Items 11 and 13 amongst these and that is in the Experimental Group as well. Items 4 and12 also show large change, more so in the Experimental Group but the change is not significant.

Table 4.8: Students who are Good in Memorization Learn History Quickly

	Control			Experimental		
	Pre Test	**Post Test**	**∆C%=(Post-Pre)%**	**Pre Test**	**Post Test**	**∆E%=(Post-Pre)%**
Strongly Disagree	2	2	0	4	4	0
%	5.0	5.0		9.3	9.3	
Disagree	6	9	7.5	11	17	13.90
%	15.0	22.5		25.6	39.5	
Neutral	7	8	2.5	8	6	4.60
%	17.5	20.0		18.6	14.0	
Agree	21	17	10	13	14	2.4
%	52.5	42.5		30.2	32.6	
Strongly Agree	4	4	0	7	2	11.60
%	10.0	10.0		16.3	4.7	
Total	40	40	20	43	43	32.50
%	100.0	100.0		100.0	100.0	

Chi - Square = 1.088	Chi - Square = 4.386
D.F = 4	D.F = 4
P-value = 0.89	P-value = 0.356
Cramer's V = 0.117	Cramer's V = 0.226

The stance of respondents is categorized into a Copier, Borrower and Criterialist Stance as described by the authors, Maggioni et al. (2004). Item 1 of BLTHQ i.e. ‘Students who are good in memorization learn history quickly,’ represents the Copier stance. The respondents overall response for this item in Control and Experimental Groups at pre-test and post-test is displayed in Table 4.8. The majority of the respondents (52.5%) in the Control Group at post-test agree with this Copier statement whereas in the Experimental Group 37.3% respondents agree with this statement of Item 1 at post-test.

The Chi Square test is used to see the difference in two sets of responses at different levels of agreement and disagreement. The Chi Square value for the Control Group is found to be 1.088, while in the Experimental Group, the value is 4.386 and shows that more difference has taken place in the Experimental Group as compared to the Control Group. The significance value of Chi Square for both sets of data shows that the difference is insignificant for this item. The result of Cramer's V for the Control Group is 0.117 showing that there is strong association between stances and groups. In the Experimental Group, the Cramer's V estimate is 0.226 and shows that there is very strong association between disagreement and post-test. This result indicates that the respondents' disagreement with this item in post-test is greater than pre-test. The Spearman rank co-relation also indicates that co-relation is stronger in the Experimental Group than the Control Group.

Table 4.9: Corroborating Evidence and Identifying Sources are Important Learning Strategies in History but Only After Mastering the Basic Facts

	Control			Experimental		
	Pre Test	**Post Test**	**∆C%=(Post-Pre)%**	**Pre Test**	**Post Test**	**∆E%=(Post-Pre)%**
Strongly Disagree	0	0		1	0	2.3
%	0	0		2.3	0	
Disagree	3	1		3	4	2.3
%	7.5	2.5	5	7.0	9.3	
Neutral	3	1		5	2	6.9
%	7.5	2.5	5	11.6	4.7	
Agree	24	26		25	22	6.9
%	60.0	65.0	5	58.1	51.2	
Strongly Agree	10	12		9	15	14
%	25.0	30.0	5	20.9	34.9	
Total	40	40	20	43	43	32.40
%	100.0	100.0		100.0	100.0	

Chi- Square = 2.262	Chi - Square = 4.120
D.F = 4	D.F = 4
P-value = 0.520	P-value = 0.0390
Cramer's V = 0.168	Cramer's V = 0.219

Table 4.9 shows the observed responses and Chi Square test for Item 2. The Chi Square test value for Control Group is 2.262 and for Experimental Group the value is 4.120. The greater value of Chi Square in Experimental Group shows the difference is greater in Experimental Group as compared to Control Group. Also, the significance value of Experimental Group is 0.0390 and shows change or difference is significant in the Experimental Group.

It is possible to calculate from Table 4.9 that most respondents, 90% in the Control Group agree with this item (I refer to Strongly Agree and Agree as a single

Agree response for interpretation). In the Experimental Group, almost the same picture emerges from the Table 4.9 that most respondents agree with the item for the Copier stance.

The total change in the Experimental Group is greater than the Control Group. In the Experimental Group, the change is 32.40% and in Control Group the change is 20%, which shows that more change has taken place in the Experimental Group as compared to the Control Group. The result in terms of total change is almost similar to Item 1 of the Copier stance.

Table 4.10: Teachers Need to Avoid Giving Students Conflicting Sources, since it Makes Historical Investigation Impossible

	Control			Experimental		
	Pre Test	Post Test	∆C%=(Post-Pre)%	Pre Test	Post Test	∆E%=(Post-Pre)%
Strongly Disagree	6	7	2.5	10	11	2.3
%	15.0	17.5		23.3	25.6	
Disagree	18	22	10	13	22	21
%	45.0	55.0		30.2	51.2	
Neutral	6	2	10	2	2	0
%	15.0	5.0		4.7	4.7	
Agree	6	7	2.5	12	5	16.3
%	15.0	17.5		27.9	11.6	
Strongly Agree	4	2	5	6	3	7
%	10.0	5.0		14.0	7.0	
Total	40	40	30	43	43	39.60
%	100.0	100.0		100.0	100.0	

Chi- Square = 3.221
D.F = 4
P-value = 0.522
Cramer's V = 0.196

Chi - Square = 6.244
D.F=4
P-value = 0.182
Cramer's V = 0.269

A sum of 60% of respondents of the Control Group disagree with this item at pre-test, and at post-test the percentage of disagreement increases to 72.5%. In the Experimental Group, 53.5% of respondents disagree with this item at pre-test, and after Intervention, the percentage of disagreement increases to 76.80%. 41.9% respondents agree with this item at pre-test and at post-test, the percentage of agreement decreases to 18.6%. These changes in responses of participants show that the Intervention in the Experimental Group as compared to

the Control Group did have some effect even if it was non-significant and that it was in the direction of a more sophisticated Criterialist stance (see Table 4.10).

The Chi Square test for the Control Group at 0.05 level of significance shows that there is no difference in responses of pre-test and post-test as its p-value is 0.522. The Experimental Group p-value of Chi Square is 0.182, which is greater than 0.05 and shows no difference in the Experimental Group at post-test as well. The Cramer's V estimate of the Experimental Group is found to be 0.2569 and shows that there is a moderate association among disagreement and post responses.

Table 4.11: Knowledge of The Historical Method is Fundamental for Historians and Students Alike

	Control			Experimental		
	Pre Test	**Post Test**	**∆C%=(Post -Pre)%**	**Pre Test**	**Post Test**	**∆E%=(Post -Pre)%**
Strongly Disagree	1	0	2.5	0	0	0
%	2.5	0		0	0	
Disagree	9	7	5	9	5	9.3
%	22.5	17.5		20.9	11.6	
Neutral	8	7	2.5	11	5	14
%	20.0	17.5		25.6	11.6	
Agree	17	13	10	15	23	18.6
%	42.5	32.5		34.9	53.5	
Strongly Agree	5	13		8	10	
%	12.5	32.5	20	18.6	23.3	4.7
Total	40	40	40	43	43	46.60
%	100.0	100.0		100.0	100.0	

Chi- Square = 1.250
D.F = 4
P-value = 0.870
Cramer's V = 0.248

Chi - Square =5.299
D.F = 4
P-value = 0.151
Cramer's V = 0.260

Table 4.11 shows observed frequencies of respondents for this Criterialist item at pre-test and post-test for both Control and Experimental Groups. A Chi Square test for both Control and Experimental Groups shows difference between pre-test and post-test responses. However, the Chi Square value for the Control Group is 1.250, and for the Experimental Group it is 5.299, which shows that more difference occurs in the Experimental Group as compared to the Control Group.

From the Table 4.11 above, it is seen that most respondents agree with the Criterialist stance. Interestingly, the Criterialist stance sees most agreement in the research of Maggioni et al. (2004) and Maggioni et al. (2009a). Here in my study, 64.0% of the respondents agree with the Criterialist stance in the Control Group while in the Experimental Group, 76.8% of the respondents agree with the Criterialist stance. In the Experimental Group, a notable change can be seen as 34.9% of the respondents agree with the Criterialist stance at pre-test and at post-test, this increases to 53.5% of the overall respondents. The Chi Square Test statistic shows that more change happened in the Experimental Group as compared to the Control Group.

This is the first item of the Criterialist stance in the Likert Scale and most show agreement, which means a sophisticated stance on the categories. 55% agree at pre-test and 65% at post-test in the Control Group, which is a difference of 10%. In the Experimental Group, 54% agree at pre-test and 77% at post-test. The percentage of difference in pre-test and post-test in the Control Group is 10% as opposed to 24% in the Experimental Group, which is more than twice.

The 1.5% difference in agreement at entry pre-test is not large between Control and Experimental Groups. The 24% difference in change between Control and Experimental Groups that is seen at post-test is very large in comparison. The p- value however, is not statistically significant.

Table 4.12: Comparing Sources and Looking for Author Subtext are Essential Components of The Process of Learning History

	Control			Experimental		
	Pre Test	**Post Test**	**∆C%=(Post-Pre)%**	**Pre Test**	**Post Test**	**∆E%=(Post-Pre)%**
Strongly Disagree	0	0	0	0	0	0
%	0	0		0	0	
Disagree	0	1	2.5	6	1	11.7
%	0	2.5		14.0	2.3	
Neutral	5	2	7.5	7	3	9.3
%	12.5	5.0		16.3	7.0	
Agree	22	19	7.5	24	20	9.3
%	55.0	47.5		55.8	46.5	
Strongly Agree	13	18		6	19	
%	32.5	45.0	12.5	14.0	44.2	30.2
Total	40	40	30	43	43	60.50
%	100.0	100.0		100.0	100.0	

Chi - Square = 3.312	Chi - Square = 12.295
D.F = 4	D.F = 4
P-value = 0.346	P-value = 0.006
Cramer's V = 0.201	Cramer's V = 0.238

This key statement of the Criterialist stance sees nearly all participants agreeing. There is a slight change in the post-test in both groups. In the Control Group 92.5% respondents agree with this item and in the Experimental Group 90.7% respondents agree with this item.

The Chi Square test and its associated p-value for the Control Group shows a difference that is insignificant between pre-test and post-test responses. In the Experimental Group, the Chi Square test statistic is 12.295, and its p-value is 0.006 and shows that there is a highly significant difference in pre-test and post-test

responses in the Experimental Group. Also, the Chi Square value for the Experimental Group is larger than that of the Control Group.

88% participants agree in the Control Group at pre-test and 92% at post-test, which is a difference of 3%. In the Experimental Group, 69% agree at pre-test and 91% at post-test, a difference of 22%, making the change significant in the Experimental Group. The difference at pre-test between Control and Experimental Groups is 18%, which is a large difference.

Table 4.13: Students need to be taught to Deal with Conflicting Evidence

	Control			Experimental		
	Pre Test	**Post Test**	**∆C%=(Post-Pre)%**	**Pre Test**	**Post Test**	**∆E%=(Post-Pre)%**
Strongly Disagree	0	0	0	1	0	2.3
%	0	0		2.3	0	
Disagree	1	0	2.5	2	1	2.4
%	2.5	0		4.7	2.3	
Neutral	3	0	7.5	2	0	4.7
%	7.5	0		4.7	0	
Agree	23	21	5	23	20	7
%	57.5	52.5		53.5	46.5	
Strongly Agree	13	19	15	15	22	16.3
%	32.5	47.5		34.9	51.2	
Total	40	40	30	43	43	32.70
%	100.0	100.0		100.0	100.0	

Chi - Square = 5.216	Chi - Square = 4.867
D.F = 4	D.F = 4
P-value = 0.157	-value = 0.301
Cramer's V = 0.255	Cramer's V = 0.238

Nearly all participants of both Control and Experimental Groups agree that students need to be taught to deal with conflicting evidence.

The Chi Square test for Control and Experimental Group shows there is no large change in Control and Experimental Groups at pre-test and post-test. However, the Chi Square value of the Control Group is larger as compared to the Experimental Group indicating that change is greater in the Control Group than the Experimental Group. The large percentages of positive responses in both groups

could indicate that the desirability of the expected response was apparent to respondents. The Intervention provided experiences with conflicting evidence as well, thus reinforcing the concept.

A 100% of participants in the Control Group agree with this Criterialist item at post-test making this change important. The Experimental Group is not far behind at 98%. The difference at pre-test is 2% between Control and Experimental Groups for this item.

Table 4.14: It is Fundamental that Students are Taught to Support their Reasoning with Evidence and Ask that History Textbook Authors do so Also

	Control			Experimental		
	Pre Test	**Post Test**	**∆C%=(Post-Pre)%**	**Pre Test**	**Post Test**	**∆E%=(Post-Pre)%**
Strongly Disagree	1	0	2.5	0	0	0
%	2.5	0		0	0	
Disagree	1	1	0	8	1	16.3
%	2.5	2.5		18.6	2.3	
Neutral	5	0	12.5	6	2	9.3
%	12.5	0		14.0	4.7	
Agree	17	21	10	15	22	16.3
%	42.5	52.5		34.9	51.2	
Strongly Agree	16	18		14	18	
%	40.0	45.0	5	32.6	41.9	9.3
Total	40	40	30	43	43	51.20
%	100.0	100.0		100.0	100.0	

Chi- Square = 6.539
D.F=4
P-value = 0.162
Cramer's V = 0.201

Chi - Square = 9.269
D.F=4
P-value = 0.026
Cramer's V = 0.238

The Chi Square test for this last question of the Criterialist stance shows different results for the Control and Experimental Groups. In the Control Group, the p-value is 0.162, which shows that there is no significant difference in the pre-test and post-test stances of respondents. In the Experimental Group, the p-value of the Chi Square test is 0.026, which reveals a significant difference in pre-test and post-test responses. Similarly, the Chi Square test statistics for Experimental Group is 9.269 and for Control Group the value is 6.539. The larger

value of Chi Square in the Experimental Group shows that the difference is larger in the Experimental Group than in the Control Group. In the Control Group, 97.5% respondents agree with this item and in the Experimental Group, 93.1% respondents agree with this item.

It is worth noting that in some items of the Criterialist stance, at pre-test, participants appear to be more naive in the Experimental Group. In this item for example, 82% agree at pre-test in the Control Group, and 68% in the Experimental Group agree that it is fundamental that students are taught to support their reasoning with evidence and ask that history textbook authors do so also. Both groups move towards more agreement at post-test but the difference in percentage is higher in the Experimental Group, which makes it significant. There is a 15% change from pre-test to post-test in the Control Group and a 26% change in the Experimental Group. This is actually a problem as the non-random nature of the sample makes chance less likely as the reason for the difference in the sample at pre-test. Here, the comparison of small groups where the population is from single cities and schools becomes important.

A reminder that the overall combined data is of 5 study groups in three cities and 3 different school systems. It can be said however, that the combined group is more representative of the larger population of history teachers in Pakistan. Another problem could be that the items have a social desirability factor at work. In the post-test especially, Intervention 1 being experienced by all, has made a difference in persuading the participants of the desirability of the pedagogical practice incorporating the historical method. That being the thrust of the workshops, it is understandable. The key question is, if it is Intervention 2 which makes for more change in the Experimental Group, or it is the effect of the difference in the sample at pre-test? Did the Experimental Group simply have more room to grow? This effect was spotted and countered in the responses to the BM questions as well. An analysis of data in the Criterialist items is additionally presented in Table 4.15 to answer the question raised above.

Table 4.15: Students need to be Aware that History is Essentially a Matter of Interpretation

	Control			Experimental		
	Pre Test	**Post Test**	**∆C%=(Post -Pre)%**	**Pre Test**	**Post Test**	**∆E%=(Post -Pre)%**
Strongly Disagree	1	0	2.5	0	0	0
%	2.5	0		0	0	
Disagree	1	1	0	8	1	16.3
%	2.5	2.5		18.6	2.3	
Neutral	5	0	12.5	6	2	9.3
%	12.5	0		14.0	4.7	
Agree	17	21	10	15	22	16.3
%	42.5	52.5		34.9	51.2	
Strongly Agree	16	18		14	18	
%	40.0	45.0	5	32.6	41.9	9.3
Total	40	40	30	43	43	51.20
%	100.0	100.0		100.0	100.0	

Chi- Square = 1.486	Chi - Square = 3.442
D.F=4	D.F=4
P-value = 0.829	P-value = 0.382
Cramer's V = 0.136	Cramer's V = 0.200

In Table 4.15 the observed frequency shows that 75% of the respondents in Control Group agree with this question at pre-test and 72.5% of the respondents agree at post-test. In the Experimental Group, 79.10% of the respondents disagree with this question with no change in pre-test and post-test response. The Chi Square test statistics for Control Group are smaller than the Experimental Group, which indicates that more difference occurs in Experimental Group than in the Control Group. Also, the Cramer's V estimate indicates that the association is moderate in the Experimental Group.

4.7.5.2 Comparison of Percentage Change and Chi Square Value for Control and Experimental Groups to Identify a Pattern in the Data

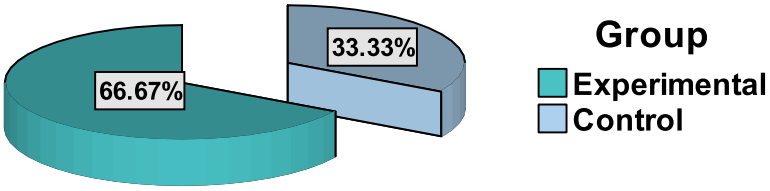
Table 4.16: Comparison of Chi Square Value and Percentage Difference in Control and Experimental Groups

Items	∆C%	χ^2 Control Group	∆E%	χ^2 Experimental Group	Comments
Q1	20	1.088	32.5	4.386	Experimental group
Q2	20	2.262	32.4	4.120	Experimental group
Q3	45	7.462	27.9	3.711	Control group
Q4	40	1.250	46.6	5.299	Experimental group
Q5	10	1.250	51.3	6.083	Experimental group
Q6	25	4.718	18.6	1.168	Control group
Q7	30	3.221	39.6	6.244	Experimental group
Q8	50	7.882	14	0.706	Control group
Q9	30	3.900	37.3	5.983	Experimental group
Q10	20	1.571	25.4	5.963	Experimental group
Q11	30	3.312	60.5	12.295	Experimental group
Q12	30	5.216	32.7	4.867	Experimental group
Q13	30	6.539	51.2	9.269	Experimental group
Q14	15	3.311	27.8	5.188	Experimental group
Q15	15	2.540	23.2	2.126	Experimental group
Q16	10	1.486	20.9	3.422	Experimental group
Q17	30	2.99	9.3	0.368	Control group
Q18	45	7.042	23.4	2.746	Control group
Q19	20	2.278	18.5	1.654	Control group
Q20	60	8.085	9.2	0.844	Control group
Q21	30	2.064	13.9	3.132	Control group

∆C%: Percentage difference from pre-test to post-test in Control Group

∆E%: Percentage difference from pre-test to post-test in Experimental Group

Figure 4.8: Comparison of Change in Control and Experimental Group



The BLTHQ analysis of all 21 items was carried out and a great deal of information was obtained through it. However, I required an overview of the observations in the 21 items to find a pattern in the data if possible. The analysis from Sections 4.7.3 to 4.7.6 presented the change or difference in responses from pre-test to post-test in Control and Experimental Groups. The change or difference in overall 21 BLTHQ items is observed in two ways. First, there is a comparison of Chi Square value in Control and Experimental Groups. Second, the absolute difference of percentage from pre-test to post-test in each category (Strongly Disagree to Agree) is taken. The Chi Square Test statistics value and their corresponding significance value are mentioned in Tables 14-B to 27-B in Appendix-B. The absolute differences of percentages in Control and Experimental Groups are also given in Tables 14-B to 27-B in Appendix-B. A closer analysis of each item shows that the Chi Square value and absolute difference of percentage are related to each other.

As the Chi Square value of a particular item in the Experimental Group is greater than the Chi Square value of the Control Group, the sum of absolute percentage difference is also greater in the Experimental Group. This larger value of Chi Square and absolute difference shows that difference or change is greater in the Experimental Group than the Control Group. On the other hand, if an item has larger Chi Square value as compared to the Experimental Group then its sum of absolute difference is also greater as compared to the Experimental Group. For example, in the analysis, Item 1 in the Control Group has a Chi Square value of 1.088 and in the Experimental Group the value is 4.386. This larger value of Chi Square in the Experimental Group indicates that the difference is greater in the

Experimental Group. Also, the sum of percentage of absolute difference in the Control Group is 20% and in the Experimental Group the sum is 32.5%. This value also indicates that change in terms of percentage is greater in the Experimental Group than the Control Group. The Chi Square value and the sum of absolute percentage difference for all 21 BLTHQ items for both Control and Experimental Groups are given in Table 4.16. The result in the Table shows that 13 items out of 21 items have a larger absolute percentage difference and Chi Square value in the Experimental Group (except Items 12 and 21, which have larger absolute difference value and smaller Chi Square value). This analysis shows that difference is consistent and greater in the Experimental Group as compared to the Control Group.

4.7.6 Effect Size to Measure the Effectiveness of Intervention

The difference in absolute percentage difference of the Control and Experimental Groups is analyzed by using independent t test. The mean of percentage difference of the Control Group is found to be 28.80 and the mean of the Experimental Group is 29.34. The significance value of test statistics, assuming equal variance; assumed as suggested by Levene's Test for Equality of Variances, is 0.901. This value reveals that there is no significant difference among percentage difference of the Control and the Experimental Group. However, as discussed earlier, as a researcher, I cannot rely only on the significance value. An effect size is calculated that measures the effectiveness of the treatment. For the calculation of effect size, I used Cohen's d effect size calculator available at http://www.uccs.edu/~lbecker/psy590/escalc3.htm. The Cohen's d effect size for the percentage difference of the Control and the Experimental Groups is found to be 0.511, in using Cohen's d: 0 - 0.20 weak effect, 0.21 - 0.50 modest effect, 0.50 - 1.00 moderate effect. Thus, from my data I can say that Intervention in my case has moderate effect.

4.7.7 Analysis of Selection of the Neutral Option

An analysis of the neutral stance taken by respondents on individual items of the BLTHQ was carried out. There are 9 Copier items, 8 Borrower items and 4 Criterialist items in the BLTHQ. A completely neutral position on a stance would require the respondent to tick a neutral on all items of the stance. No one took a neutral stance on all items in both pre-test and post-test. Two respondents in the Control Group at pre-test took a neutral stance on all 4 Criterialist items, and one respondent in the Experimental Group took a completely neutral stance on all 8 items of the Borrower stance at post-test, which is probably not a meaningful result.

What is perhaps more interesting is the result of the neutral position in the Criterialist items. 16.3% of all 160 options on Criterialist items were ticked at neutral in the Experimental Group at pre-test and this reduced to 5.8% at post-test. This result was almost identical in the Control Group, which is probably due to the fact that at pre-test, the numbers in both Control and Experimental Groups were identical. There were 14.7% of all 360 neutral options taken at pre-test for Copier items, which reduced to 10.8% at post-test in the Control Group. There was a similar difference in the Experimental Group on this going from 12.7% to 8.5%. In Borrower items, 21.6% of all 320 options were selected at neutral and this changed to 14.7% in the Control Group. In the Experimental Group the change was from 13.4% to 12.2 % which is negligible. As there were 43 respondents in the Experimental Group the number of options is slightly higher than in the Control Group with its 40 participants.

Thus, only the Criterialist stance in both Control and Experimental Groups showed a large change of almost a third, less selections of the neutral option at post-test. The other bit of interesting information is that in the Borrower items, there were 21.6% options ticked at neutral at pre-test, which reduced to 14.7% at post-test in the Control Group. The difference in the Experimental Group however, is only of 1% which is strange as the difference in the Control Group is of 7% at post-test.

Generally, however, there is a clear pattern of fewer options of neutral being ticked at post-test as compared to the pre-test. This can be attributed to the experience of Intervention 1 while the effect of Intervention 2 is not seen as a difference in the Experimental Group as far as the neutral stance is concerned.

Table 4.17: Neutral Option Analysis in the Control and Experimental Group

	Control				Experimental			
	Pre	**%**	**Post**	**%**	**Pre**	**%**	**Post**	**%**
Copier	46	12.78	39.00	10.83	43	11.11	31	8.01
Criterialist	21	13.13	9.00	5.63	21	12.21	10	5.81
Borrower	64	20.00	47.00	14.69	44	12.79	34.	9.88

4.7.8 Analysis of Individual Change from Pre-Test to Post-Test

In the analysis of change in stance of the respondents as a comparison of pre-test to post-test responses, the frequencies show the number of respondents who chose a change from, say, an Agree towards a Disagree, or vice versa, in Copier, Borrower, or Criterialist stance.

The neutral box indicates where respondents chose a neutral more times than at pre-test. The No Change box tells the number of respondents that did not change in terms of numbers at agreement or disagreement.

The earlier analysis of all items of the BLTHQ by Contingency tables showed that respondents disagreed with most of the Copier items. However, the Contingency tables display the responses to overall items and is not a calculation of change of stance of individual respondents. For a closer analysis of individual change, the Agree and Strongly Agree positions were first collapsed as agreement. Similarly, Disagree and Strongly Disagree positions were merged into a single Disagree position. The position of agreement and disagreement of individual respondent is then analyzed. The method is described below.

For example, a respondent agrees with 3, disagrees with 4 items and takes a neutral on 2 items of the Copier stance at pre-test. In the post-test he/she agrees with 5 items, disagrees with 2 items and selects a neutral on 2 items of the Copier stance. In this case, he/she appears to agree more with Copier items and, on the other hand, he/she appears to disagree less with the Copier items at post-test. So his/her stance is calculated to be more agreement with Copier items than disagreement with Copier items at post-test. Table 4.18 of the Control Group and Table 4.19 of the Experimental Group display overall change of respondents obtained by summing all the respondents with change in stance obtained as described above. The Tables thus display frequencies and percentages of change. In the Control Group, with the Copier stance, 50% of the respondents change towards agreement while 22.5% change towards disagreement, 15% change towards neutral and 12.5% remain at their original position. In the Criterialist items, 32.5% change towards agreement, 12.5% change towards disagreement, 5% change towards neutral position, and 50% remain at their entry level. 32.5% of the respondents change towards agreement in the Borrower items, 50% change towards disagreement, 15% change towards neutral position and only 2.5% show no change from pre-test to post-test data.

In the Experimental Group, the picture which emerges from the data is quite different and interesting in terms of changes. In the Copier items, almost 50% of the respondents change towards disagreement, which is completely against and in the opposite direction as compared to the Control Group, 20.9% change towards agreement, 11.6% change towards neutral and 18.6% of the respondents remain at their entry level with no change in their stance from pre-test to post-test. In the Criterialist items, 35.7% change towards agreement and 33.3% change towards agreement. Here in the Experimental Group, change towards disagreement in the Criterialist items increased as compared to the Control Group where only 12.5% changed towards disagreement. 21.4% of the respondents show no change in their stance, which is very low as compared to the Control Group, where 50% do not change their position in the Criterialist items. In the Borrower items, 65.1% change

towards agreement, 9.3% change towards disagreement, and 25.6% have no change in their stance.

Figure 4.9: Comparison of Pre-Test to Post-Test Change of Copier, Criterialist and Borrower Stance in Control Group

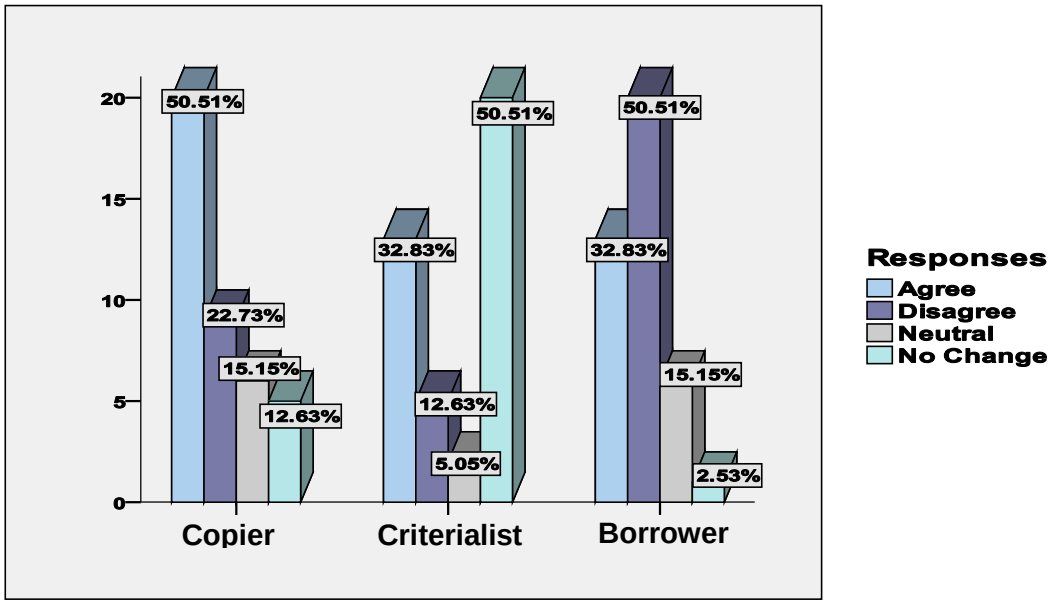


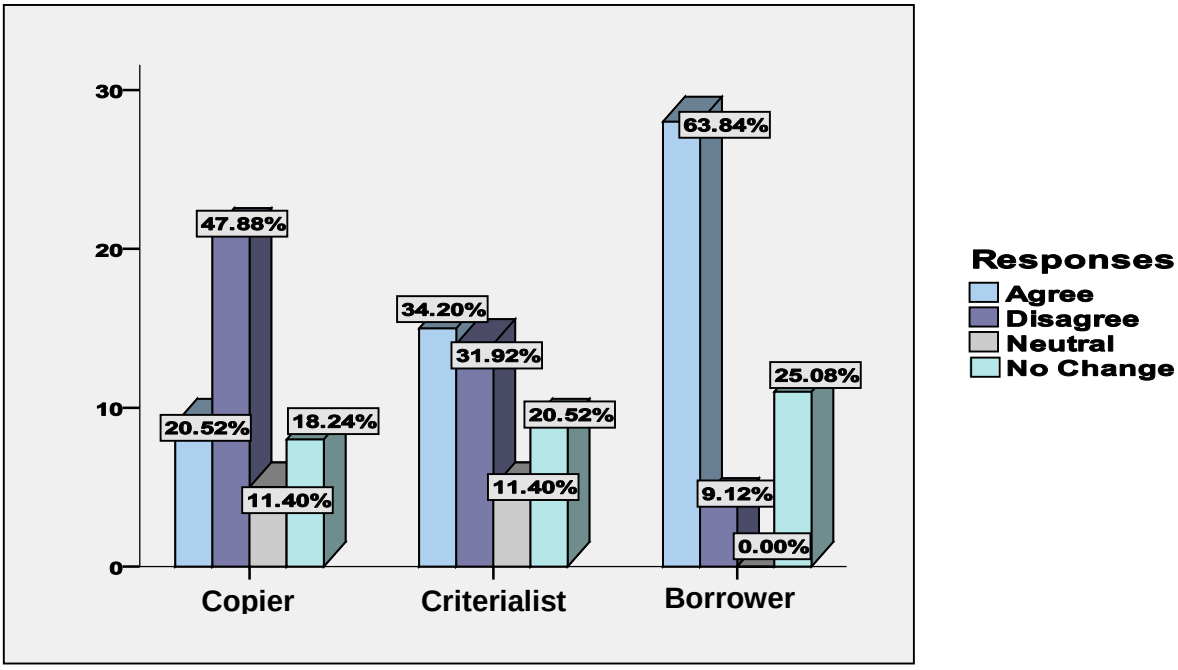
Table 4.18: Analysis of Change in the Control Group

Control			
Stance	**Copier**	**Criterialist**	**Borrower**
Agree	20	13	13
	50.0%	32.5%	32.5%
Disagree	9	5	20
	22.5%	12.5%	50.0%
Neutral	6	2	6
	15.0%	5.0%	15.0%
No change	5	20	1
	12.5%	50.0%	2.5%

Table 4.19: Analysis of Change in the Experimental Group

Stance	**Experimental**		
	Copier	**Criterialist**	**Borrower**
Agree	9	15	28
	20.9%	35.7%	65.1%
Disagree	21	14	4
	48.8%	33.3%	9.3%
Neutral	5	5	0
	11.6%	11.9%	.0%
No change	8	9	11
	18.6%	21.4%	25.6%

Figure 4.10: Comparison of Pre-Test to Post-Test Change of Copier, Criterialist and Borrower Stance in Experimental Group



4.7.9 Analysis of Pre-Test to Post-Test Difference in Control and Experimental Groups

4.7.9.1 Wilcoxen Rank Sum Test

Table 4.20: Results of the Wilcoxen Rank Sum Test in Control and Experimental Groups

Items	Control Group		Experimental Group	
	Z Statistics	Significance Value	Z Statistics	Significance Value
Q1	-0.971	0.332	-1.852	0.064
Q2	-1.396	0.163	-1.325	0.185
Q3	-1.413	0.158	-1.409	0.159
Q4	-2.253	0.024	-2.182	0.029
Q5	-0.275	0.783	-1.079	0.280
Q6	-1.194	0.233	-0.821	0.412
Q7	-1.413	0.158	-2.375	0.018
Q8	-0.280	0.779	-0.436	0.663
Q9	-0.864	0.387	0.000	1.000
Q10	-0.857	0.392	-2.066	0.039
Q11	-1.126	0.260	-3.590	0.000
Q12	-2.524	0.012	-1.794	0.073
Q13	-1.451	0.147	-2.623	0.009
Q14	-1.110	0.267	-0.626	0.531
Q15	-0.936	0.349	-1.424	0.154
Q16	-0.034	0.973	-0.784	0.433
Q17	-1.153	0.249	-0.492	0.623
Q18	-0.165	0.869	-0.827	0.408
Q19	-0.646	0.518	-0.861	0.390
Q20	-0.996	0.319	-0.647	0.517
Q21	-1.330	0.183	-0.544	0.586

The difference from pre-test to post-test responses in the Experimental Group is explored in the previous analysis. However, to find the consistency of change in the Experimental Group, another statistical research technique is used. The data of the BLTHQ is obtained on a Likert Scale and possesses an ordinal nature (Strongly Disagree to Strongly Agree). Therefore, instead of using a paired t test to measure the significance of difference in both sets of data, a non-parametric Wilcoxen Rank Sum Test is used to measure the difference in pre-test and post-test responses in both sets of data. Table 4.20 gives the result of the Wilcoxen Rank Sum Test statistics with their corresponding significance value in Control and Experimental Groups.

In the Control Group, the difference from pre-test to post-test responses can be seen as significant in Item 4 and Item 12 (significance value is highlighted with red color in Table 4.20). This shows that the stance of respondents in pre-test is not significantly different to the post-test in all items except Item 4 and Item 12. In the Experimental Group, 5 items: Q4, Q7, Q10, Q11 and Q10 (significance value with green color in Table 4.20) show a significant difference in responses. Also, Q1 (0.064) and Q12 (0.073) have very small significance value and can be considered significant. Overall 7 items out of 21 items show significant difference in their responses and the significance value of the remaining questions are also smaller in the Experimental Group than the Control Group. This indicates that the Intervention in the Experimental Group may possibly have effected change in beliefs of respondents thus showing more change taking place in the Experimental Group. The Wilcoxen Rank Sum Test analysis above is compatible with the results of Chi Square and percentage change analysis, which shows that more change occurred in the Experimental Group than in the Control Group.

Table 4.21: Analysis of Pre-Test to Post-Test Responses in Control and Experimental Group in a Single Study (Rose Group)

	Control		Experimental	
Items	**Z-statistics**	**Significance Value**	**Z-statistics**	**Significance Value**
Q1	-0.707	0.480	-1.633	0.102
Q2	-0.431	0.666	-1.134	0.257
Q3	-0.142	0.887	-.033	0.041
Q4	-0.707	0.480	-1.342	0.030
Q5	-0.520	0.603	-0.378	0.705
Q6	-0.736	0.461	-0.378	0.705
Q7	-1.633	0.102	-0.730	0.465
Q8	-0.647	0.518	-1.890	0.049
Q9	-0.816	0.414	-0.378	0.705
Q10	-2.232	0.026	-1.000	0.317
Q11	-1.265	0.206	-1.841	0.036
Q12	-1.342	0.180	-0.277	0.028
Q13	-0.577	0.564	-1.838	0.021
Q14	0.000	1.000	-1.134	0.047
Q15	-0.791	0.429	-1.414	0.157
Q16	-10.394	0.163	-0.577	0.564
Q17	-1.633	0.102	-1.089	0.016
Q18	-0.368	0.713	-.0447	0.655
Q19	-1.903	0.057	-.0447	0.655
Q20	-2.050	0.040	-0.272	0.785
Q21	0.000	1.000	0.000	1.000

The effect of Intervention 2 in the Experimental Group is seen in the meta-analysis of overall data of the BM. Differences in small groups support the hypothesis as more change in the Experimental Groups in 4 out of 5 studies. All groups except Pansy show change in responses as more significant in the Experimental Group compared to the Control Group. A similar trend can also be seen in the BLTHQ and EBI data where result of meta-analysis indicates that more items have significant change in the Experimental Group as compared to items of the Control Group. The number of items being 21 in the BLTHQ Likert Scale and 32 in the EBI, a study wise analysis of difference was difficult. The numbers in cells were small in the single studies as well as unequal in numbers of participants. One

single study was analyzed by means of the same statistical test used for overall data of 5 studies. Though the sample size is smaller than overall data, the result is quite consistent and meaningful. The result indicates that only two items in the Control Group have significant difference in pre-test to post-test responses, while in the Experimental Group, 8 items including all 4 Criterialist items show significant difference of pre-test to post-test responses of respondents. See Table 4.21.

4.7.10 Analysis of Change and No Change in Control and Experimental Groups

All the above analysis consistently indicates that there is change in both groups and the change in respondents' stance is greater in the Experimental Group than in the Control Group. To examine the pattern of this difference or change, another analysis is carried out. The overall data in both Control and Experimental Groups are categorized into Change and No Change categories.

I assign a category of change for all of those respondents who shift from their original position from pre-test to post-test, and assign No Change for all of those respondents who remained at their original position from pre-test to post-test. Thus, the whole data is classified into Change and No Change categories. Now the likelihood of change in the Experimental Group is compared to No Change for each item of the BLTHQ by means of Odds Ratio. Table 4.22 shows the Odds Ratio of all 21 items with their corresponding confidence intervals. Table 4.22 shows that the Odds Ratio of 10 items is greater than 1, and 2 items have Odds Ratio close to 1. This indicates that in most BLTHQ items, the Experimental Group exhibits more change than the Control Group. I have interpreted Item 2 for the reader to understand the logic of Odds Ratio in my study. The Odds Ratio for Item 2 is found to be 1.263, which is greater than 1. This means that for each respondent in the Experimental Group, change is 1.263 times more likely than No Change. The Odds Ratio of the remaining items can be interpreted in the same manner.

More importantly, 3 out of 4 Criterialist items show a high Odds Ratio of more change taking place in the Experimental Group. The Criterialist stance is the

stance of higher sophistication on the scale. As the authors argue, consistency in epistemic beliefs is seen in reference to the Criterialist stance.

Table 4.22: Odds Ratio (OR) with 95% Confidence Interval (CI) of BLTHQ Items

Items	Odds Ratio	95% CI	If OR > 1 then Experimental Group is better than Control Group
Q1	0.580	0.243__1.386	Each respondent in Experimental Group is less likely to be changed
Q2	1.263	0.532 __2.997	Each respondent in Experimental Group is more likely to be changed
Q3	0.748	0.308__1.818	Each respondent in Experimental Group is less likely to be changed
Q4	0.774	0.326__1.841	Each respondent in Experimental Group is less likely to be changed
Q5	1.535	0.645__3.654	Each respondent in Experimental Group is more likely to be changed
Q6	1.544	0.649__3.671	Each respondent in Experimental Group is more likely to be changed
Q7	0.934	0.392__2.26	Each respondent in Experimental Group is less likely to be changed

Q8	0.514	0.213__1.243	Each respondent in Experimental Group is less likely to be changed
Q9	1.556	0.654__3.703	Each respondent in Experimental Group is more likely to be changed
Q10	0.787	0.332__1.864	Each respondent in Experimental Group is less likely to be changed
Q11	1.381	0.574__3.321	Each respondent in Experimental Group is more likely to be changed
Q12	1.895	0.791__4.537	Each respondent in Experimental Group is more likely to be changed
Q13	5.714	2.040__16.007	Each respondent in Experimental Group is more likely to be changed
Q14	0.926	0.386__2.223	Each respondent in Experimental Group is less likely to be changed
Q15	1.396	0.588__3.314	Each respondent in Experimental Group is more likely to be changed
Q16	1.143	0.481__2.713	Each respondent in Experimental Group is more likely to be changed
Q17	0.592	0.245__1.431	Each respondent in Experimental Group is less likely to be changed

Q18	0.800	0.318__2.012	Each respondent in Experimental Group is less likely to be changed
Q19	0.554	0.227__1.352	Each respondent in Experimental Group is less likely to be changed
Q20	0.541	0.219__1.339	Each respondent in Experimental Group is less likely to be changed
Q21	1.150	0.486__2.723	Each respondent in Experimental Group is more likely to be changed

Red indicates where the respondent is more likely to change in the Experimental Group. The green indicates a high score in favour of the Experimental Group, which is, however, less than 1.

4.7.11 Exploratory Factor Analysis of the BLTHQ Items (Appendix-C)

Exploratory Factor Analysis, with the principal component Analysis Extraction Method, was performed separately on data from the BLTHQ items of both Control and Experimental Groups. One objective of Factor Analysis in my study was to see if the factors that emerge in my analysis represent a single Copier, Borrower or Criterialist stance or appear to be representing more than one stance. The authors of the BLTHQ have carried out the same study and describe Factor Analysis on both sets of data (i.e. pre-test and post-test). They also consider a two factor solution in their study, as suggested by the Scree Plot and name the two factors with the help of negative or positive loadings. Factors that emerge in the authors' analysis consist of almost all items of Copier, Criterialist and Borrower stance with different positive and negative loadings. I carried out the Factor

Analysis to identify the pattern in my data whether the factors that emerge consist of items of different stance or appear to be a single Copier, Criterialist or Borrower stance.

In my study, using the same instrument, the BLTHQ, Q 1, Q 2, Q 3, Q 5 to Q 10 represent the Copier stance. Q 4, Q 11, Q 12, Q 13 represent the Criterialist stance. Q 14 to Q 21 represent the Borrower stance.

The KMO Bartlett test of sphericity in pre-test responses of Control Group data is 0.396, which is very low, and below the standard requirement. However, in post-test data of the Control Group, the KMO test value is 0.507, which is fairly good compared to the post-test. Therefore, I ignored these smaller values and carried out Factor Analysis on both sets of data.

The post-test data shows more certain and consistent behaviour and the reliability is fairly good (see Table 4.23). In the Control Group, the reliability scale ranges from 0.680 to 0.820 except the reliability of Factor 4, which is 0.436. In the Experimental Group, values range from 0.577 to 0.807 which show high consistency in the data.

In the Control Group, eight factors emerged in the pre-test with Eigen values greater than one, which is not encouraging. However, I considered the first four components as important factors for all the sets of data because they explain most variation in all sets of data. The pre-test for both Groups, Control and Experimental, show random and uncertain behavior of respondents as seen in an earlier item analysis. The reliability of the factors at pre-test in both sets of data is low. In the Control Group, the reliability value ranges from 0.387 to 0.645 for all four factors. In the Experimental Group at pre-test, the reliability scale ranged from 0.154 to 0.579 showing a poor consistency in the data.

I next considered the test of the Control and Experimental Groups using the BLTHQ as 4 different sets of test in order to see a difference in the way Factors emerged in the samples. I carried out Factor Analysis on both sets of data, however, for explanation and discussion I considered only the post-test result of Control and Experimental Groups. For interpretation and naming of factors that emerged from the data, I calculated the threshold or cut off value of all four factors

and considered only those items in the factors whose loading is greater than their cut off value and ignored all of those whose loading is less than their cut off value regardless of their negative and positive sign.

Overall, there is a change in Factor loading at post-test in both Control and Experimental Groups with a more definable pattern in the Experimental Group.

The change in post-test in both Control and Experimental Groups may be taken as consistent with expectations as the research design incorporates two Interventions. Intervention one is provided to both Control and Experimental Groups, hence the change in both groups at post-test is understandable. Intervention two was provided to the Experimental Group only and the data suggests a difference in the degree and quality of change in stance between Control and Experimental Groups. The difference in the Control and Experimental Groups at post-test can be judged to reflect this difference in the experience provided to both. This, interestingly, is also consistent with the picture that emerged with the data from the other instrument, the Boscolo & Mason questions. The difference in change between the two groups is significant and cannot be ignored.

It is important to note that Change and Degree of Change is important in both instruments at post-test as described above. The change is apparent and the direction of change or Degree of change in the beliefs targeted by the BLTHQ and the BM instruments appears similar. The nature of the change, from a Borrower stance to a more Criterialist stance, and from a Copier stance to a more Criterialist stance that is observed in data from the BM analysis can be compared to change as seen in the responses to the BLTHQ. This is interesting in the light of the difference of the questions or statements in both instruments. The BM asks theoretical epistemic questions whereas the BLTHQ nests questions related to epistemic beliefs in contexts of pedagogical practice. The BLTHQ questions are set in the context of teaching and learning history and relate to the teachers' practice more directly. This is discussed in the overall analysis. (The items are argued by the authors to be reasonable proxies for the epistemic beliefs named Copier, Borrower or Criterialist). The change, as seen in the BM analysis, suggests

a move towards a more Criterialist stance whereas the change as observed in the BLTHQ also suggests a move towards a more Criterialist stance. The difference in the statements of the BLTHQ and BM are used to discuss the responses in more depth.

Table 4.23: Reliability of All Factors in the Control and Experimental Groups

Control post-test			Experimental post-test	
Factors	**Cronbach α**	**No of Item**	**Cronbach α**	**No of Item**
Factor 1	0.778	5	0.806	3
Factor 2	0.720	7	0.695	7
Factor 3	0.721	7	0.602	5
Factor 4	0.337	4	0.695	4
Factor 5	0.484	3	0.192	4
Factor 6	0.679	2	0.303	7
Factor 7	0.497	2	0.501	2
Factor 8	-	-	0.258	5

Table 4.24: Principal Component Analysis (PCA) of Post-test Questionnaires of Control Group

Factors	Linear Combination	Cut off value	% of Variation	Reliability (Cronbach α)
Factor 1	0.87Q21 + 0.76(Q12 + Q13) + 0.6Q4 + 0.4Q11	0.4	21.72	0.778
Factor 2	0.8Q9 + 0.7(Q8+Q6) + 0.6Q3 - 0.4(Q1 + Q5 - Q11) + 0.3(Q4 - Q13)	0.4	15.89	0.720
Factor 3	0.9Q19 + 0.6(Q5 + Q10 + Q18) + 0.5Q3 + 0.4(Q1 + Q7+ Q14)	0.5	9.719	0.721
Factor 4	0.8Q15 + 0.7Q17 + 0.5Q18 - 0.4(Q10 + Q1 + Q4)	0.4	7.99	0.337

Table 4.25: Principal Component Analysis (PCA) of Post-test Questionnaires of Experimental Group

Factors	Linear Combination	Cut off value	% of Variation	Reliability (Cronbach α)
Factor 1	0.9Q13 + 0.8Q12 + 0.711	0.7	17.590	0.806
Factor 2	0.8Q1 + 0.7Q6 + 0.5Q21 + 0.4 (Q17 + Q7 + Q2 + Q3)	0.4	13.57	0.733
Factor 3	0.8Q8 + 0.7Q19 + 0.6Q7 + 0.5Q6 - 0.4(Q12+Q21)	0.4	12.02	0.577
Factor 4	0.8Q15 + 0.7Q17 + 0.6Q16 + 0.4Q18	0.4	7.05	0.610

4.7.12 Factor Analysis of the BLTHQ Combined Data (166 Respondents)

In the earlier part of the study, I carried out Factor Analysis of data in separate sets of pre-test and post-test in both the Control and the Experimental Groups. The factors that emerged in each set of data were compatible with the other statistical analysis in the sense that at pre-test, factors did not appear coherent but the position improved at post-test. Data of the Experimental Group showed a clear Criterialist Factor with high reliability. The second factor that emerged was a Copier with a single item of the Borrower stance.

I decided to run Factor Analysis of the combined data of the Control and the Experimental Groups, pre-test and post-test, to see if a different picture emerged.

In this final Factor Analysis using the complete set of data, the Scree Plot (Appendix-C, Figure 1C) and Eigen value criteria suggest that seven factors again emerged from the data. I considered 0.3 as a cut off or threshold value for all the factors and considered only those items whose loading was greater than 0.3 regardless of the negative and positive signs. Factor 1 explains 15.86% of variance and appears to be completely a Copier factor with a single Borrower item, Item 19. All Copier items appear with high to low loading with high loading on Item 5, Item 6, Item 1, moderate loading on Item 7, Item 8, Item 9, and low loading on Item 3. Reliability of Factor 1 is measured by Cronbach alpha and it is found to be 0.735, which I can take as fairly good. Factor 2 explains 13.214% of the variance and it appears to be a completely Borrower Factor with a single Copier item with very low loading. Factor 2 contains Items 14, 15, 17 and 18 with moderate loading, and Item 19 with low loading. The reliability of Factor 2 was found to be 0.622. Factor 3 explains 7.6% of the total variance and it appears with all Criterialist items with high loading on Items 12 and 13, moderate loading on Item 11, and low loading on Item 4. Factor 3 completely represents a Criterialist stance. Factor 3 might be considered important although Factor 3 explains only 7.6% of the total variance. I consider that the percentage of variance is less important. What is important is that Factor 3 exclusively represents the Criterialist stance. Reliability of Factor 3 was found to be 0.424 but after dropping Item 4 in the analysis, the reliability value

increased to 0.658. These first three Factors are considered important as they appear to be of a single stance in each case.

The remaining four factors contain items loading together from various stances. These are also explained below.

Factor 4 explains 6.5% of the total variance with negative moderate loading on Item 9 of the Copier stance, positive high loading on Item 16, and positive low loading on Item 15 of the Borrower stance. So this factor represents contradiction of the Copier and Borrower stance. Factor 5 explains 5.9% of the variance and appears with high positive loading on Item 4 of the Criterialist, and high positive loading on Item 20 of the Borrower stance. I name this factor as a combination of Criterialist and Borrower stance. Factor 6 explains 5.08% of the variance and appears with high positive loading on Item 21 of the Borrower stance, moderate positive loading on Item 10 and Item 3 of the Copier stance. Therefore, I name this factor a combination of Copier and Borrower items. Factor 7 explains 4.82% of the variance and it represents Item 1, Item 2 and Item 3 of the Criterialist stance.

4.7.13 Weighted Scores and Consistency Scores of Epistemic Stance for All Participants

Table 4.26: Student Weighted Averages Scores on Epistemic Stances and Consistency Scores of Control Group

Respondent Number	Copier		Criterialist		Borrower		Consistency Score	
	Pre	Post	Pre	Post	Pre	Post	Pre	Post
1	-0.78	-0.78	1.5	2	0.75	0	66.67	71.43
2	-1.22	-0.56	2	2	0.75	-0.75	61.11	75.00
3	-0.33	-0.22	0.75	1.5	0.5	0.75	57.89	55.00
4	0.00	0.44	0.5	1.25	0	0.125	57.14	50.00
5	-0.33	-0.22	1	0.5	0.75	0.75	50.00	50.00
6	-0.22	-0.56	1.75	0.75	-0.375	-0.75	61.11	57.89
7	-0.22	0.00	1.25	1.5	0.875	0.375	50.00	52.63
8	0.00	0.56	1.5	1.25	1.5	1.125	40.00	40.00
9	-0.44	-0.89	1.75	1.75	0.875	0.375	57.89	73.68
10	-0.89	-0.56	0.5	0	-0.625	-0.75	77.78	78.95
11	0.11	0.44	1.75	1.5	0.375	0.125	47.62	57.14
12	-0.33	-0.67	1.25	0.5	-0.375	-0.25	78.57	77.78
13	-0.89	-0.56	1.5	1.5	0.75	0.25	63.16	61.11
14	-0.67	-1.11	1	1.5	0.375	-0.125	62.50	78.95
15	0.11	-0.33	1	1.75	0.625	0.875	47.37	50.00
16	0.00	0.11	0.5	0.75	0.625	1	46.15	35.00
17	0.33	0.67	1	1.75	0.75	1.375	38.10	38.10
18	-1.33	-1.56	1.75	2	0.875	1.25	71.43	61.90
19	0.22	0.44	0.5	1	0.375	0.75	41.67	31.25
20	-0.33	-0.78	0.5	1.25	0.125	0.5	53.33	60.00
21	0.44	0.67	1.75	2	0.625	0.625	47.37	37.50
22	-1.33	-0.44	0.5	0.75	0.75	0.375	63.16	60.00

Respondent Number	Copier		Criterialist		Borrower		Consistency Score	
	Pre	Post	Pre	Post	Pre	Post	Pre	Post
23	0.11	0.22	0.75	1.5	0.625	0.375	44.44	50.00
24	-0.22	-0.44	0.25	0.75	0.25	-0.375	50.00	73.68
25	0.56	0.67	1	1.75	0.5	1.375	38.89	28.57
26	0.11	-0.22	1.25	1	0.875	0.375	47.06	58.82
27	0.00	0.11	0	1	0.875	0.625	38.89	43.75
28	0.44	1.33	1.5	2	1.125	1.25	33.33	20.00
29	0.11	-0.33	0.75	0.5	0.25	-0.125	46.67	66.67
30	0.00	0.33	-0.75	1.75	0.5	1.25	33.33	46.15
31	-0.78	-0.22	0.25	1.25	0.125	0.75	76.92	55.00
32	0.22	0.11	-0.25	1	0.25	0.125	38.46	57.14
33	-0.56	-0.89	1	1.25	0.75	1.125	62.50	56.25
34	-0.11	1.56	1.25	0.75	0.25	0.125	57.14	42.86
35	0.44	0.33	0	0.75	0	0.125	20.00	50.00
36	0.11	-0.78	1.25	1	0.625	0	45.00	76.19
37	-0.11	-0.22	0.75	0.75	0.125	0.25	61.54	61.54
38	-1.44	-1.44	1	2	0.125	0.5	61.90	70.00
39	-0.44	-1.33	1.75	2	0.5	0.625	61.90	71.43
40	-0.22	0.22	1.5	0.5	0	0.5	62.50	55.56
Median	-0.11	-0.22	1.00	1.25	0.50	0.38	53.3	56.695
S.D	0.5184	0.69	0.62	0.54	0.42	0.56	12.91	14.55

The median value of the Copier stance at pre-test in the Control Group is -0.11, which shows that 50% of the respondents have weighted average less than -0.11. The remaining 50% of the respondents have weighted average greater than -0.11. At post-test, the median value shows more disagreement towards the Copier stance. The median value of the Consistency score at pre-test and post-test in the Control Group shows that most of the respondents have low consistency scores. On the other hand, the median consistency score of the Experimental Group at

post-test is 61.9%, which shows that almost 50% of the respondents have less than 62% consistency score, and the remaining 50% have moderate and high scores.

Table 4.27: Student Weighted Average Scores on Epistemic Stance and Consistency Scores of the Experimental Group

Respondent Number	Copier		Criterialist		Borrower		Consistency Score	
	Pre	Post	Pre	Post	Pre	Post	Pre	Post
1	-0.78	-0.67	0.75	0.5	0.375	1	66.67	58.82
2	-0.56	-0.22	0.25	1.75	0.75	0.25	46.15	68.75
3	-0.67	0.22	0.5	1.25	0.5	0.5	57.14	50.00
4	0.33	-0.56	1	1.75	0.125	0.25	50.00	68.42
5	-1.44	-0.89	0.25	1	0.25	0.5	61.90	66.67
6	-0.33	0.11	0.75	1.25	0.875	0.625	52.38	45.00
7	-0.56	0.00	-0.5	1	-0.5	-0.5	65.00	71.43
8	-1.44	-0.89	2	2	0.875	0.875	70.00	61.90
9	-0.44	0.11	1	1.5	0.5	0.25	56.25	55.56
10	0.00	-0.11	1	1	-0.125	-0.25	57.14	68.42
11	-0.78	-1.11	0	1	-0.625	-0.125	63.16	76.19
12	-1.00	-0.78	0.75	0.25	0.25	0.375	71.43	45.00
13	-0.22	-0.56	0.75	1.25	0.375	1	45.00	52.38
14	-0.11	-0.22	1.25	1.25	-0.375	0.75	71.43	55.00
15	0.56	0.22	0.25	1.5	0.625	0.625	42.86	50.00
16	0.56	-0.44	-0.5	1	-0.125	0.625	35.29	56.25
17	0.78	-0.11	1	-0.25	0.5	1.5	31.25	35.29
18	-0.67	-0.33	1	1.5	0.875	0.75	57.89	60.00
19	-1.11	-1.22	1.25	1.5	0	0	71.43	76.19
20	0.22	0.67	0.5	1	0.125	0.625	46.67	29.41

Respondent Number	Copier		Criterialist		Borrower		Consistency Score	
	Pre	Post	Pre	Post	Pre	Post	Pre	Post
21	-0.78	-0.89	1	1.25	0	0.625	71.43	66.67
22	0.33	-0.22	0.75	1.25	-0.25	-0.125	52.94	63.16
23	0.67	0.44	0.75	1.25	0.625	-0.125	42.86	52.38
24	0.00	-0.33	0.5	1	0.125	0.125	60.00	65.00
25	-0.22	-0.33	1.25	1.5	0.875	0.25	55.00	57.89
26	-1.11	-0.89	0.75	0.5	0.125	-0.125	65.00	75.00
27	-0.44	-0.22	1.25	1	0.25	0.375	68.75	60.00
28	-0.22	-0.56	1.5	2	0.125	0.25	60.00	83.33
29	0.11	-0.11	1	1.25	0.625	0.875	44.44	45.00
30	-0.89	-1.11	1.5	2	0.625	0.125	66.67	75.00
31	0.00	-0.22	1	1.25	1	0.875	47.37	57.14
32	0.56	0.22	1	1	0.875	0.625	29.41	42.11
33	-0.89	-1.00	1.25	1	0.25	-0.375	68.42	85.00
34	-1.00	-1.11	0.75	1	0.5	0.125	68.75	75.00
35	-0.11	-1.11	1	2	1.625	1.5	38.89	61.90
36	-0.11	-1.11	1	1.5	0.625	1	60.00	68.42
37	0.11	-1.33	0.5	1.75	-0.75	0.125	65.00	75.00
38	-0.22	-0.89	1.25	2	-0.125	-0.75	65.00	78.95
39	-0.78	-0.67	0	0.5	-0.25	-0.375	57.89	65.00
40	0.67	1.00	0.75	1.5	0.625	1.125	59.09	33.33
41	-0.44	-1.00	0	1	0	-0.375	63.16	90.48
42	0.33	0.33	1.5	1.25	1.375	1	36.84	45.00
43	0.11	0.00	0.5	2	0.25	0.5	50.00	61.11
Median	-0.22	-0.33	0.75	1.25	0.25	0.375	57.89	61.9
S.D	0.59	0.56	0.52	0.49	0.50	0.53	11.70	14.02

To obtain the weighted average, I did some complex calculation following those carried out by Maggioni et al. (2004). For the weighted average of each

respondent, I assigned a +2 to Strongly Agree, +1 to Agree, 0 to Neutral, -1 to Disagree,

and -2 to Strongly Disagree. I then made a sum of the scores on all items and divided the sum by the number of items of that stance. For example, respondent 1 at pre-test in the Control Group agreed with 1, disagreed with 6 and strongly disagreed with 1 Copier item. So the weighted score for the Copier stance is [(1x1+(-1x6) + (-2x1)]/9. The weighted average for the respondent in Copier items is -0.78. The weighted average for the Criterialist items is calculated in the same fashion and it is found to be 1.5. For the Borrower stance the weighted average is 0.75. If the weighted score of a respondent in Copier items is a positive +1, it shows the agreement of the respondent towards Copier items. If a respondent obtained a weighted average of +1, +2, +3 in the Criterialist stance and -1, -2, -3 in the Copier and Borrower stance, then the respondent is considered to be more of a Criterialist.

The weighted average is calculated to ascertain the magnitude of change from pre-test to post-test. Whereas, in order to assess how consistent is the stance, a consistency score of each respondent is calculated. The consistency score is calculated in relation to the Criterialist stance as the stance of higher sophistication and the direction towards which the teaching was directed. VanSledright and Reddy (2014) set the bar for consistency at 90%, which is high in their own estimation. Maggioni and VanSledright also report (2014, p.12)[41] an average 68% for students and 74-75% for teachers. I take into account that this was the first time that participants had been introduced to epistemic ideas of beliefs in knowledge in history and that the length of the Intervention was no more than 3 days spread over three weeks for most participants. This was also probably among the first few Likert Scale questionnaires that respondents had ever filled, and certainly the first time they had encountered questions such as these relating to epistemic beliefs, whereas the respondents in the authors' study would be more familiar with the territory. Therefore, I have set a lower bar at 60%, moderate at 70% and high at 80%.

[41] This is a Paper presented at AERA 2014 awaiting publication.

Maggioni et al. (2004) found that respondents appeared to agree with the Criterialist stance as well as the Borrower or other stance to some extent. The data in my study is most discomfiting in this regard. Generally, over 90% agree with the Criterialist stance in both groups at post-test but nearly 60% agree with the Borrower stance and 40% agree with the Copier stance at the same time. There are various ways in which this result can be reasoned.

There is a common pattern in the four tests, pre-test and post-test and Control and Experimental Groups as well as within the 5 small groups, of more agreement with the Criterialist stance, some agreement and larger disagreement with the Copier stance and some concurrent agreement with the Borrower stance. There is consistence as well in the assessment of more change in the Experimental Group. The direction of change from pre-test to post-test is also generally towards greater sophistication as was the thrust of the teaching in the workshops. This common pattern points towards what could possibly be a successful assessment of stance and change in stance by the BLTHQ.

It may be argued that the participants are just generally confused in epistemic understanding more so in the pre-test. There is low to moderate development in sophistication at post-test but not to a satisfactory level if seen as groups rather than individually. There is, however, a difference in Control and Experimental Groups and within small groups.

Finally, a consideration of whether the criteria that underpins an assumption that practice choices are associated by the teachers as reflecting epistemic beliefs and therefore, needing to be coherent can be questioned. The teachers did select the Criterialist option as desirable but also other options, therefore they did not make the link that agreement with one should mean a disagreement with the other. If it is necessary to identify a stance for respondents, then some decisions need to be taken. The choice is to ignore agreement with the Criterialist stance as it appears to be a common decision that is desirable and focus on the responses, indicating the Copier and Borrower stance and the difference within these from pre-test to post-test.

If this is done, the results make sense. An examination of the median suggests that there is disagreement with a Copier stance in both groups, pre-test and post-test, and more so in the Experimental Group at post-test. The pre-test indicates more difference in the Control and Experimental Groups in the Borrower stance. The median in the Control Group is 0.50 at pre-test and reduces by half to 0.25 in the Experimental Group. At post-test the Control Group median reduces to 0.38 and increases to 0.38 in the Experimental Group. A closer examination of change in small groups is seen as necessary to understand this.

Table 4.28: Student Consistency Scores on Epistemic Stances of Control Group and Experimental Group

Student Number	Control Group		Experimental Group	
	Pre Test	Post Test	Pre Test	Post Test
1	66.67	71.43	66.67	58.82
2	61.11	75.00	46.15	68.75
3	57.89	55.00	57.14	50.00
4	57.14	50.00	50.00	68.42
5	50.00	50.00	61.90	66.67
6	61.11	57.89	52.38	45.00
7	50.00	52.63	65.00	71.43
8	40.00	40.00	70.00	61.90
9	57.89	73.68	56.25	55.56
10	77.78	78.95	57.14	68.42
11	47.62	57.14	63.16	76.19
12	78.57	77.78	71.43	45.00
13	63.16	61.11	45.00	52.38
14	62.50	78.95	71.43	55.00
15	47.37	50.00	42.86	50.00
16	46.15	35.00	35.29	56.25
17	38.10	38.10	31.25	35.29
18	71.43	61.90	57.89	60.00
19	41.67	31.25	71.43	76.19
20	53.33	60.00	46.67	29.41
21	47.37	37.50	71.43	66.67
22	63.16	60.00	52.94	63.16
23	44.44	50.00	42.86	52.38
24	50.00	73.68	60.00	65.00
25	38.89	28.57	55.00	57.89

Student Number	Control Group		Experimental Group	
	Pre Test	Post Test	Pre Test	Post Test
26	47.06	58.82	65.00	75.00
27	38.89	43.75	68.75	60.00
28	33.33	20.00	60.00	83.33
29	46.67	66.67	44.44	45.00
30	33.33	46.15	66.67	75.00
31	76.92	55.00	47.37	57.14
32	38.46	57.14	29.41	42.11
33	62.50	56.25	68.42	85.00
34	57.14	42.86	68.75	75.00
35	20.00	50.00	38.89	61.90
36	45.00	76.19	60.00	68.42
37	61.54	61.54	65.00	75.00
38	61.90	70.00	65.00	78.95
39	61.90	71.43	57.89	65.00
40	62.50	55.56	59.09	33.33
41	-	-	63.16	90.48
42	-	-	36.84	45.00
43	-	-	50.00	61.11
Mean	53.01225	55.923	56.18488	61.22209
Median	51.665	56.695	57.89	61.9
Minimum	20	20	29.41	29.41
Maximum	78.57	78.95	71.43	90.48
Range	58.57	58.95	42.02	61.07

Following the method used by Maggioni et al. (2004), the consistency score of each respondent in the Control and Experimental Groups is calculated by comparing agreement of the respondent to the Criterialist stance. I assigned 0 to all respondents who agree to Copier and Borrower items and thus disagree with

Criterialist items. I assigned 1 to all of those respondents who disagree with Copier and Borrower items and agree with Criterialist items. I then sum all the respondents at 1 and divide by the total numbers of items that have been responded to by each individual. For example, if a respondent agrees with Items 1, 2, 3, 5 and disagrees with 6, 7, 8, 9, 10, he/she has 4 zeros and 5 ones. Similarly, if he/she agrees with all Criterialist items he/she has 4 ones. If he/she agrees with 16, 18, 21, and disagrees with 14, 15, 17, 19, 20, he/she has 3 zeros and 5 ones. So, the consistency score of this respondent is $\frac{(5+4+5)}{21}$ = 0.666 or 66.6%. I consider those respondents who have a consistency score greater than 80% at post-test to be consistent with the Criterialist stance.

The Criterialist stance is the stance of greater sophistication in the scale of Copier, Borrower, and Criterialist. It is also the stance towards which the Intervention targeted change in epistemic concepts therefore it is reasonable to assess consistency against the Criterialist stance as suggested by Maggioni et al. (2004). The consistency scores are next ordered as Low, Moderate, and High Scores according to Table 4.29.

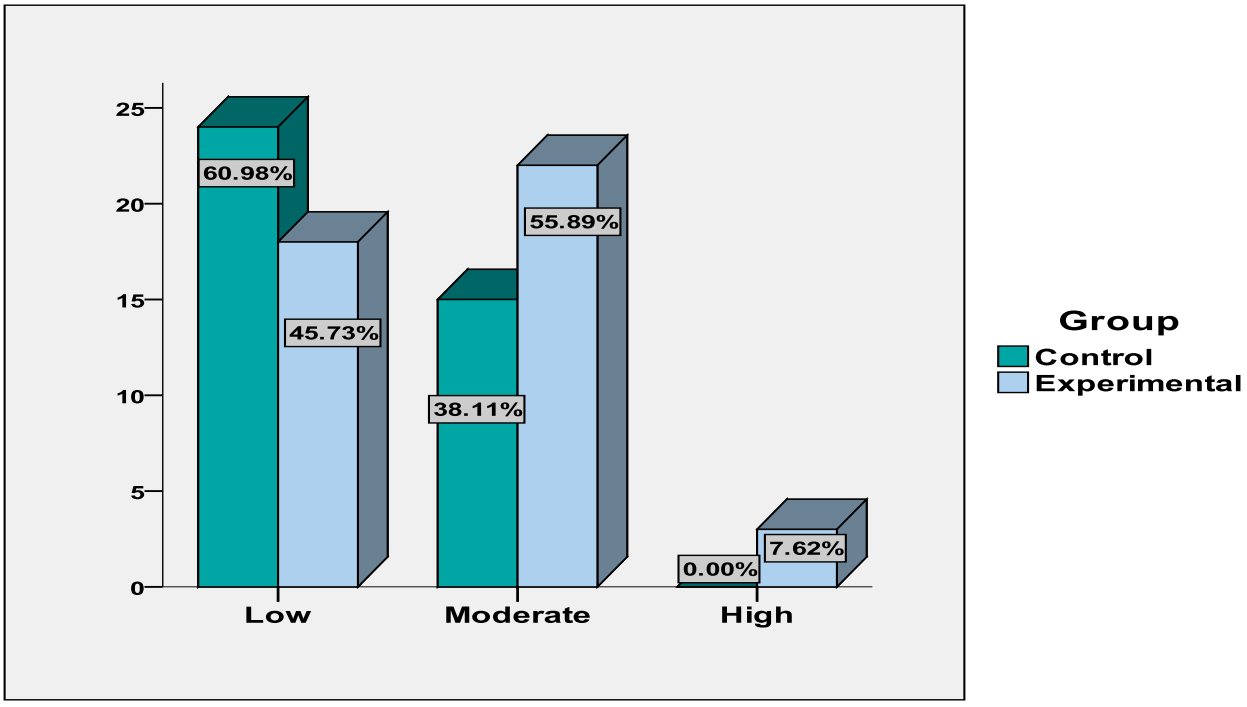
4.7.14 Weighted Averages and Consistency Score for the Control and Experimental Group

Weighted averages and consistency scores of individual respondents in the 5 individual studies were also calculated in the above mentioned procedure. The results indicate that there is no clear pattern of significant change observed in small groups as seen in the overall combined data. This is therefore not included in the analysis. All tests are included in the Appendices.

Table 4.29: Categories of Low, Moderate and High Consistency Score of Epistemic Beliefs

Consistency Score	Range	Control Group		Experimental Group	
		Pre	Post	Pre	Post
Low	Below 60%	25	24	23	18
%		62.5	60	53.48	41.86
Moderate	60% to	15	16	20	22
%	79%	37.5	40	46.5	51.16
High	80% to	0	0	0	3
%	100%	0	0	0	6.9

Figure 4.11: Comparison of Consistency Score of Control and Experimental Group



I consider those respondents, who have a score at 80 or above, to have high consistency with the Criterialist stance; those with a score ranging between,

60-79 to have moderate consistency, and those below 60 to have low consistency with the Criterialist stance. Taking a percentage according to the numbers in each test i.e. 40 respondents in the Control Group and 43 in the Experimental Group, I can develop an overview. According to this classification of my data, I can say that there were 40% at moderate consistency with the Criterialist stance in the Control Group at post-test, and 60% at low consistency with the Criterialist stance at post-test.

In the Experimental Group, there were 53% at low consistency with the Criterialist stance at pre-test and this changed to 42% at post-test. There were 47% at moderate consistency and this changed to 51% at moderate consistency in the post-test in the Experimental Group. There were 7% at high consistency in the Experimental Group alone at post-test.

The difference in the weighted averages of Copier, Criterialist and Borrower stance is estimated by using a paired t test. In the Control Group only, the weighted average of the Criterialist stance shows significant difference from pre-test to post-test significance value 0.009 (p-value). The mean of the pre-Criterialist stance is 0.97 and the mean of the post-Criterialist stance is 1.26, which shows that the difference is towards the post-test.

The Criterialist stance shows a highly significant difference from pre-test to post-test in the Experimental Group. The mean of the pre-test Criterialist stance is 0.79 and the post-test Criterialist stance mean is 1.25, which is a high weighted score in post-test.

4.7.15 Paired t Test of Pre-Test and Post-Test Consistency Scores in Control and Experimental Groups

The paired t test is also used among pre-test and post-test consistency scores of Control and Experimental Groups. The result indicates that only in the Experimental Group, the consistency score of the pre-test is significantly different from the post-test significance value 0.0029 (p-value), while in the Control Group,

no significant difference occurred in pre-test to post-test consistency scores (p-value 0.548).

4.8 Epistemic Beliefs Inventory

4.8.1 Summary and Discussion of Analysis of Data of the Third Measure, the Epistemic Beliefs Inventory

The EBI, Epistemic Belief Inventory, is a measure of general Epistemic Beliefs. For the Analysis of the EBI, the purpose of analysis is mainly to identify change from pre-test to post-test in both sets of data, Control and Experimental. I need to study if the pattern of consistently more change in the Experimental Group seen in the BLTHQ and BM analysis, is found in the data of domain general epistemic beliefs as measured by the EBI. Results are not significant but interesting. The pattern is supported as more difference of the difference in the Experimental Group in 20 out of 32 items in the EBI. This pattern is more pronounced in the EBI than in the BLTHQ data and is consistent with the findings of the two instruments.

A Chi Square test of independence was taken for each item. However, Chi Square values were not significant in most items. A pattern could be seen on the Contingency Tables with respondents moving in some degree and direction. This needed to be examined in some detail. Therefore, the Chi Square analysis was followed by an analysis of a difference of the difference in observed responses from pre-test to post-test in the Control and Experimental Groups separately. The analysis shows that there is somewhat more difference in responses from pre-test to post-test in terms of percentages in the Experimental Group as compared to the Control Group.

An independent t test was carried out to find if the difference in percentage difference of the Control and Experimental Groups is significant. The test failed to reveal any significant difference amongst the two groups. An effect size was also calculated to measure the effectiveness of the intervention in the Experimental

Group. The Cohen d effect size for percentage difference shows moderate effect of the Intervention in the Experimental Group. More importantly, an overview of difference in all 32 items shows *consistent* more difference in the Experimental Group as compared to the Control Group. This *consistent* difference supports the results as seen in the data in the other two instruments, the BLTHQ and BM. I take this as most encouraging and have presented the Tables 4.30 and 4.31 for examination.

A weighted averages analysis of the EBI was carried out. The result of the weighted averages appears different to the analysis of the difference of the difference. It must be noted that the absolute difference of each item deals with overall absolute change in stance of all respondents for a particular item (e.g. Item 1) from pre-test to post-test. Weighted averages deal with change in individual respondents within different categories of items e.g. Simple Knowledge (SK), Certain Knowledge (CK), Omniscient Authority (OA), and median stance is taken for comparison. De Vaus (2001, p.153) argues that distinguishing between aggregate level and individual change is important both at the theoretical level and at the level of research design. The aggregate change will not tell us about the level of individual change.

The change from pre-test to post-test is seen in a sum of all subscales, Quick Learning, Fixed Ability, Simple Knowledge, Certain Knowledge and Omniscient Authority. This analysis further examined the magnitude of change from pre-test to post-test in three subscales, Simple Knowledge, Certain Knowledge and Omniscient Authority only, as these are of interest to me with reference to epistemic beliefs. Median points show no pre-test and post-test change was observed in SK and OA stance in both groups, Control and Experimental. There is a small difference of degree of disagreement on knowledge as certain between Control (-0.57) and Experimental Groups (-0.13) at post-test. In case of CK, Certain Knowledge, respondents tend to disagree with CK in both groups at pre-test and post-test with slightly more disagreement at post workshop in the Control Group. It appears that whatever difference is seen in the overall analysis has more to do with subscales, Quick Learning and Fixed Ability.

Table 4.30: If Two People are arguing about something, at least one of them must be wrong

	Control			Experimental		
	Pre Test	**Post Test**	**∆C%=(Post -Pre)%**	**Pre Test**	**Post Test**	**∆E%=(Post -Pre)%**
Strongly Disagree	10	7	7.5	9	10	2.4
%	25.0%	17.5%		20.9%	23.3 %	
Disagree	7	10	7.5	14	16	4.6
%	17.5%	25.0%		32.6%	37.2 %	
Neutral	13	9	10	10	4	14
%	32.5%	22.5%		23.3%	9.3%	
Agree	6	7	2.5	6	7	2.3
%	15.0%	17.5%		14.0%	16.3 %	
Strongly Agree	4	7	7.5	4	6	4.7
%	10.0%	17.5%		9.3%	14.0 %	
Total %	40	40	35	43	43	28

Table 4.31: Some People Just Have a Knack for Learning and Others Don't

	Control			Experimental		
	Pre Test	**Post Test**	**∆C%=(Post-Pre)%**	**Pre Test**	**Post Test**	**∆E%=(Post-Pre)%**
Strongly Disagree	6	4	5	5	8	7
%	15.0%	10.0%		11.6%	18.6%	
Disagree	9	3	15	8	13	11.6
%	22.5%	7.5%		18.6%	30.2%	
Neutral	12	13	2.5	12	7	9.3
%	30.0%	32.5%		25.6%	16.3%	
Agree	9	15	15	10	10	0
%	22.5%	37.5%		23.3%	23.3%	
Strongly Agree	4	4	0	8	5	7
%	10.0%	10.0%		18.6%	11.6%	
Total %	40	40	37.5	43	43	34.9

Table 4.32: Percentage Difference of Items in the Control and Experimental Group

Items	Control ∆C%	Experimental ∆E%
1	25	23.1
2	36.4	27.9
3	36.4	**51.1**
4	20	**32.6**
5	5	**18.4**
6	15	**37.1**
7	20	**32.6**
8	15	**37.2**
9	30	**32.5**
10	30	**46.6**
11	20	**46.4**
12	10	**27.8**
13	35	**41.9**
14	30	**32.5**
15	45	27.8
16	50	41.8
17	37.5	34.9
18	27.5	21
19	35	28
20	30	18.6
21	30	**69.9**
22	40	14
23	45	23.3

Items	Control ∆C%	Experimental ∆E%
24	25	**55.7**
25	20	**27.8**
26	25	**18.6**
27	40	14
28	35	23.1
29	25	**41.7**
30	25	**37.2**
31	25	**37.3**
32	30	**37.1**

∆C%: Absolute percentage difference from pre-test to post-test in the Control Group

∆E%: Absolute percentage difference from pre-test to post-test in the Experimental Group

Bold highlights a pattern of consistent more difference in the Experimental Group as compared to the Control Group. A moderate effect is seen with a calculation of effect size although the difference is not significant.

4.8.2 Factor Analysis of the Overall EBI Data

The authors of the EBI describe the emergence of four factors in their study. Initially, the reliability scale of EBI items was very poor and it was impossible to carry out Factor Analysis in the data. 12 items were then dropped one by one by checking their reliability value. The reliability of the remaining items was found to be 0.742, which could be seen as fairly good in the context. Factor Analysis was carried out with Varimax Rotation and seven factors emerged from the data that have Eigen values greater than 1. A Scree Plot suggested that the first three factors were important with a large gap between Factors 1 and 3. The first and second Factors represent a combination of Simple Knowledge, Certain Knowledge

and Fixed Ability, of which two are epistemic beliefs and one is a belief about learning. Factor 3 explains 7.3% of the variance and contains items of Quick Learning, Simple Knowledge and one item of Omniscient Authority. This loading together of factors to do with epistemic beliefs and those related to learning is not coherent in my understanding.

Tables 4.30 and 4.31 are attached to show samples of 2 items. Table 4.32, exploring the difference of the difference in change in the Control and Experimental Groups, is also shown above.

4.9 Summary of Results of All Three Measures

Results of tests using the three instruments, the EBI, BLTHQ and the Boscolo & Mason questions are reported and discussed in this Chapter.

Qualitative assessment of the 166 responses to open ended Boscolo and Mason questions was carried out to the CBKH rubric developed for this purpose. Qualitative descriptors and quantitative scores were awarded to each respondent at pre-test and post-test. Inter-Rater and Internal Evaluation was conducted and is reported.

An overall assessment of data finds significant change in participants' stance from pre-test to post-test in both Control and Experimental Groups. The Chi square statistic, however, is different at 23.73 for the Control Group and 35.51 for the Experimental Group. Odds of change is 1.259.

The difference in the change from pre-test to post-test between Control and Experimental Groups is more evident in fine grain analysis. In an analysis of Change and No Change, the mean rank for the No Change category in the Control Group is 6.65. For the Experimental Group, the mean rank is 13.72. Significance value is 0.004. A comparison of change from low to high stance between Control and Experimental Groups is also interesting. Most participants (72.5% in the C and 72.09% in E) are at a low stance at pre-test in both groups. At post-test, 67.5% in the Control Group and 81.4% in the Experimental Group are at a high stance. Chi

Square value is 12.83 for Control and 24.81 for Experimental and the significance value is 0.000.

A comparison of data of the 5 studies finds positive effects of Intervention 2 in 4 of 5 studies. The results of this analysis of data from the BM measure are seen to support the hypothesis. The BLTHQ Likert Scale has 21 items and data is examined as an overall meta-analysis of all 5 studies as numbers in cells in individual studies is small. A coherent pattern is seen in cross tables of individual items. A Table (4.16) is produced for an overview of Chi Square values and absolute percentage difference in Control and Experimental Groups for all 21 items. Chi square values and percentage difference are seen to relate to each other. In 13 of 21 items, a larger percentage difference and smaller Chi square value in the Experimental Group is seen. Change is significant from pre-test to post-test in two items of the Control Group and 7 items of the Experimental Group. The significance value of the remaining items is also noted as being smaller in the Experimental Groups. The result of a Wilcoxen Rank Sum Test is compatible with the Chi square statistic. A comparison of change and No Change is conducted on each item. Odds ratio of 10 items is greater than 1, and 2 items have Odds ratio close to 1. For each respondent, change is more likely in the Experimental Group in 10 plus 2 items.

An analysis of percentage difference does not find significant difference between Control and Experimental Groups but effect size is 0.511, which is a modest effect according to Cohen d. Due to the number of items and the smallness of size of groups in individual studies, separate analysis of all 5 studies is not conducted. The BLTHQ data of one single study is analysed for comparison and a similar result is seen. 2 items of the Control Group and 8 items of the Experimental Group show significant change. Thus, the results are seen to support the findings of the BM data analysis.

The pattern of the data of the EBI Likert Scale shows that the result is consistent with that of the BM and BLTHQ and more change is seen in the Experimental Group. Results are not significant but the pattern of the data and results consistently favor the Experimental Group in terms of change. A Chi square

test of independence followed by percentage analysis of difference in difference in all 32 items shows that more difference has taken place in responses of respondents in the Experimental Group. Cohen d effect size was also calculated to measure the effectiveness of the Intervention in the Experimental Group. The result indicates that the Intervention has had a moderate effect in the Experimental Group.

Independent studies were not analysed separately due to the number of items, 32, and the size of the sample. Instead, data was combined in meta-analysis.

CHAPTER 5

CONCLUSION

‘The teacher’s role is more, you can say, it depends upon the teacher’s role, and his personality. Is it, is authoritarian, hmm or, or….. it is a more collaborative approach soo……..it depends how much liberty teacher is giving, yes’.

‘When liberty was given, when liberty was given when different opinions came out it.....’

- Participants of Experimental Group giving reasons for change.

5.1 Introduction

In this last Chapter, I conclude from the theoretical arguments that are explored in the research, and from the empirical evidence, to draw conclusions on the alternative hypothesis proposed. The question, if a perception of agency relates to change in concepts in the context of task based discourse in classrooms, is answered in the light of the strengths and limitations of the process. The strength of the research lies in the design to observe a relationship between two cognitive constructs, a perception of agency and epistemic belief change, in valid contexts of classroom learning and practitioner research, in the effort made to counter threats to inference with better design, open, detailed, and transparent reporting, and most importantly, the observation of more and better change in 4 of 5 studies.

What is known is thus, as argued in the review of literature, what is known to work in 4 of 5 studies, and how it is known is made available for review and interpretation. Truth claims are necessarily tentative, open to interpretation by readers. It would, however, be hard to offer a single alternative explanation for the consistency of change seen in all tests, in all three measures, and importantly, in 4 of 5 studies other than the limitations of quasi experimental design or data and statistical analysis.

This is an interesting puzzle. To explain, I, as the researcher and tutor in all 10 Control and Experimental Group workshops, would need to identify a single alternative cause that could explain the difference of more and better change in the 5 Experimental Groups. Features such as better examples or materials, more time, better feedback, etcetera could explain a difference in learning but these would have to be consistent in all Experimental Groups and missing in Control Groups. The consistency required, precludes accidental differences having an effect in 4 of 5 groups. The probability of 4 of 5 studies producing positive results by chance without a common effective cause, is perhaps not high. This is consistent in all tests and in all three measures. The difference is observed in larger numbers in change as opposed to No Change, degree change, and in assessment of change from low to high stance. There is also somewhat more coherence in the performance as seen in Factor Analysis at post-test. In the examination of change and No change and difference in low and high stance at the level of independent studies, more significant change is seen in the Experimental Groups. A D.I.D estimate finds positive results in favour of the Experimental Groups in 4 of 5 studies. Meehl (1990, p.204) argues of course, that in research in the social sciences and arguably in the biological sciences, ‘everything co-relates to some extent with everything else’. I do not argue with Meehl but present this practitioner research to be examined in its parameters of design, constraints and opportunities in educational contexts.

This finding in analysis of the BM data in 5 studies is supported in the meta-analysis of overall BLTHQ data. The analysis was difficult to carry out in each of the 5 independent studies. As there are 21 items in the BLTHQ and 32 in the EBI, and small, uneven numbers in cells, it was not attempted. BLTHQ data of one study (Rose) that has positive results in the BM data was analysed. 2 items in the Control Group and 8 items in the Experimental Group showed significant change supporting the result seen in the BM data. See Table 4.21.

The alternative hypothesis of a relationship of a perception of agency and change in concepts is thus seen as supported. This is discussed in the following

pages. Theoretical aspects are recalled briefly, questions are restated followed by a summary of results. Finally, conclusions are drawn and argued.

The perception of agency as well as epistemic beliefs, being psychological constructs, limit the possibilities of definition and the reliability of the measurement. The communication of the research is, therefore, detailed, transparent, and conclusions are tentative and open to interpretation. Having said that, it is important to recognise that such limitations and constraints will always be part of the equation when researching these constructs in the field. As discussed in the review of literature, practitioner research, quasi experiments with non-random assignment, do not permit the same kind of inference as true experiments. This relationship could not be observed in a laboratory setting due to the manipulations that would be required in sample selection and assignment making a perception of agency hard to argue.

5.2 The Argument

The research focus, a relationship between a perception of agency and change in concepts in task based discourse, is drawn from a theoretical argument. The processes of paradigm change in scientific communities as described by Kuhn are argued as different from the orchestrated learning situations of teacher education classrooms. To restate, change processes, as argued by Kuhn, are autonomous, meaningful to participants, highly charged, whilst classroom learning situations are planned, structured, may be teacher or group led, and constrain the learner's agency to reflect upon own prior knowledge and develop own concepts.[42] From this theoretical argument, I draw a hypothesis of a relationship of a perception of agency with conceptual change in task based discourse in teacher education classrooms. The relationship is logical and arguable, and the research finds between good to moderate empirical support for the alternative hypothesis.

The constructs of agency and a perception of agency are employed in this study but not interchangeably. I argue agency as a variable affordance in social

[42] 1. Thomas Kuhn. 2. See Review of Literature: An Argument for Agency

contexts on the basis of the difference I perceive in the planned and imposed process of change as produced in classrooms and Kuhn's descriptions of naturally evolving paradigm change in scientific communities. For the purpose of this research, I hypothesize an enhanced perception of agency in task based group discourse, which may extend possibilities for participants to reflect upon their own prior ideas, to question, agree or disagree with group views in order to develop their own concepts. Questions can be raised if the perception in itself was adequate for an individual to change or would a perception created in the whole group work to enhance the agency of individuals in real terms and thus make a difference? More qualitative research could help to further understand the cause of the change in stance seen in the data.

Task based, group learning situations provide the necessary contexts, actions, and opportunities for focused, purposeful talk for the development of meaning.[43] For the individual to learn, to change a concept such as epistemic beliefs, within this social context, the individual is the agent. In my understanding, educators need to examine the approach to task based group discourse to include an awareness of the agency of the individual learner within the social context. Contexts, actions, and words provide vital opportunities for talk as well as active learning and as Schön (1963) argues, and Newman (1999) agrees, convergence of meaning is made possible. Newman, however, disagrees with Schön's thesis of reflection in action, arguing instead, for language games with reference to Wittgenstein. In this research, I examine if making discourse productive for more learners requires building a perception of agency within the group discourse to question, to agree or disagree and reflect upon own prior concepts as well as the new ideas on offer. There is logic in this as it would be hard indeed to argue that concepts are changed in an environment that imposes ideas by force.

[43] See Schön (1987a, pp.100-101) in Newman (1999, p.39) as discussed in The Review of Literature.

5.3 Reflection and Language Games

A review of literature for this research finds general understanding that conceptual change is a rational, psychological process as well as a social process. Intention and motivation are also seen as important by theorists. Thus, both internal and external factors are said to play a role in conceptual change. This is also the underpinning position for my empirical research that examines change in concepts as supported by a perception of agency within task based discourse. A perception of agency, it is argued, supports participants to reflect upon own prior concepts within active, social situations to change and develop new concepts. In this study, evidence of change is seen to support this relationship albeit tentatively. An analysis of a post session focus group with participants also sheds light on thoughtful deliberation as participants talk about the change or difference in their stance after the learning experience.

On the basis of my observations, I consider that reflection is necessary and did take place in the experimental workshops. The context of task based discourse involving the use of actions and words was a necessary part of the process of learning. Practically, the participants took part in specific tasks of examining relevant and suitable evidence, making of knowledge claims, learning meaning and words such as sources, evidence, significance, as well as the process of using heuristics to assess knowledge claims. They came to recognize and learn norms of knowledge construction in the community of historians and, as the excerpt illustrates, they could later reflect upon their own naiveté with some amusement. Newman argues for convergence of meaning taking place in participation in the language games of historians and a reinterpreted reflection may, according to him, be part of the process but only ‘after at least some convergence of meaning has been achieved’ Newman (1999, p.157).

In my study, while the social aspect is seen as necessary, it is problematised with a changing and changeable perception of agency within discourse examined as a meaningful factor in change. Change is posited as a mindful, individual process of reflection within task based discourse to develop

concepts. I find tentative support for the hypothesis that a perception of agency within task based discourse facilitated reflection upon own prior concepts and those of others in developing conceptual understanding.

Skemp's theory of experience of a collection of suitable examples supporting change of concepts is also tested and found to be effective. The experience of a range of suitable examples as an effective pedagogical strategy finds significant support in my research. Thus, words, actions, experience in social contexts, participating in what can be called the language games of historians, and reflection in action were all essential parts of the process.

With an enhanced perception of agency in task based discourse, participants were able to reflect upon their prior ideas about knowledge in history, be critical, compare these with the new ideas of knowledge in history as co-constructed, and change. It is possible that participants were able to concurrently reflect upon their own prior concepts and find new ideas plausible to change to a different epistemic perspective. Change does take place all the time and perhaps this is how it happens. More research is needed to answer the question of how and why change took place.

5.4 Effects of Intervention One, Experience of a Range of Suitable Examples on Change in Concepts

Effects of Intervention One, experience of a range of suitable examples provided in matched content and method to all 10 groups of both Control and Experimental is examined. This was the method of teaching and it had to be necessarily, a carefully crafted and matched experience in order to support valid conclusions for Intervention 2, a perception of agency relating to change in concepts. Unfortunately, a Control Group that did not experience a range of suitable examples, could not be provided without adding to confusion. Interestingly, this Intervention has had excellent effects. Change is significant in both Control and Experimental Groups at post-test. I conclude from this statistic that 'experience of a range of suitable examples' does support change in concepts.

5.5 Support in the Data for the Hypothesis of a Relationship of a Perception of Agency in Task Based Discourse with Change in Concepts

Empirical evidence provides reasonable support to the hypothesis as I draw conclusions from the rather copious amounts of qualitative and numerical data. The results are tantalizing and interesting but not conclusive in the sense of statistical significance in most tests of the relationship. Effectiveness of the intervention of a perception of agency is examined with data from the three measures, the BM questions and BLTHQ and EBI Likert Scales.

More and better change in Experimental Groups is observed with moderate effect in the data of all three measures, the BLTHQ as well as the BM and the EBI while being qualitatively and quantitatively different in each. Some tests show significance. The change is consistent and more apparent in fine grain analysis and with more support in the nature and quality of change. The participants generally moved from either a more objectivist or relativist conception of knowledge towards a range of awareness of the use of evidence and argument in the choice between explanations in history. Conclusions are drawn from an analysis of overall change as Change and No Change, a difference in low and high stance at post-test, and differences in degree of change. Analysis is conducted as a meta-analysis of data of all 5 studies as well as of independent studies in varying ways appropriate to the data.

Cohen d effect size[44] ranges between 0.511 - 0.60 in BLTHQ data tests, which is moderate by Cohen's standards. Cohen et al. (2007, p.521) Cohen considers a differential measure of effect size as more important than significance Cohen et al. (2007, p.521). A Difference of Difference estimate with BM data is 0.275. This is the positive effect of the treatment.

Effect size rankings in teacher education research are not available. There is no effect size quoted in Hattie Rankings (September 2014, May 2015) on a perception of agency as this may be the first time that this relationship has been suggested, researched, or reported, to my knowledge. The sizes quoted for related

[44] Cohen (1988) argues an effect size of 1.0 as ‘ large, blatantly obvious’ Hattie (2009:7-8)

psychological constructs such as 0.48 for motivation, 0.24 affective attributes, can be taken as rough reference. Some tests show significant difference. Without an appropriate reference for research in adult or teacher education, I consider the effect sizes seen as adequate as both a perception of agency and epistemic beliefs are psychological constructs.

There is consistently more and better change seen in the Experimental Groups in nearly all tests and all three measures. This consistency in all tests of various relationships, I consider important even if statistical significance is seen in only a few tests. Effect sizes, Odds Ratio and the Difference of the Difference are positive.

In overall analysis of BM data, respondents from both Control and Experimental Groups change significantly from pre-test to post-test but Chi Square values for the Experimental Groups (35.51) are higher compared to the Control Groups (23.73). The Odds Ratio is 1.259; likelihood of more change in the Experimental Groups. A paired t test has similar significant change in both groups with higher values in the Experimental Group.

A Wilcoxen rank sum test to see the significance of the difference in change between groups was conducted. The difference in the Control Groups is significant in only 1 study of 5 while it is significant in 4 studies of 5 in the Experimental Groups.

A Mann Whitney Test of Change, No Change and Degree Change of the difference between Control and Experimental Groups is significant and reliable. The mean rank in Degree Change is 8.33 in the Control Group and 18.53 in the Experimental Group, which is a large difference between groups.

The difference in change between the Control and Experimental Groups is most apparent in the comparison of change from low to high stance. While both Control and Experimental Groups have a very similar stance at pre-test, at post-test the difference is large. 67.5% in the Control Group are at a high stance and 81.4% are at a high stance in the Experimental Group. The significance value is 0.00, which is a good result to have!

A D-I-D estimate of treatment effect is 0.275 which is positive. Odds calculated are 1.259 in favour of change in the Experimental Group with a 95% confidence interval. If the confidence interval is lowered in the context of this educational research, is seen to have more value. The notion that the alternative hypothesis is moderately supported gains ground with these results from the data of the Boscolo and Mason questions.

Fine grain analysis of Low and High stance between the 5 study groups using a D.I.D. estimate was conducted. This was positive in 4 groups and negative in 1 group. Thus, 4 Experimental Groups out of 5 show positive effect of the Intervention of a perception of agency in task based discourse.

It is interesting to note the similarity in the two groups, at pre-test. The percentage of low scores at pre-test is 72.5 in the Control Groups, and 72 in Experimental Groups. There were 27.5 in the high scores at pre-test of the Control Group and 27.9 in the Experimental Group. The total change to high scores was very different at post-test between Control and Experimental Groups. 81.4 % were at high scores in the Experimental Group as compared to 67% in the Control Group. A change can easily be seen in terms of observed counts of Low and High category in both sets of data.

In the data of the BLTHQ, 21 items were categorized into the 3 stances, the Copier, Borrower and Criterialist stance. Cross tables analysis does not show the data as random or confused with clear patterns emerging in the consistency of responses falling in easily discernible bands for most items in both groups. A Chi square test of independence (see para 4.7.5.1) supports more positive change in the Experimental Group. The BLTHQ, being a Likert Scale, each item is analysed. In order to observe a pattern a table was made. Chi Square values, percentage difference and Odds Ratio were displayed and compared for each item. Thus, it is possible to see that 13 out of 21 items show more positive change in the Experimental Groups compared to the Control Groups. Cohen's d effect size 0.511 of the percentage difference in Control and Experimental Groups is a moderate effect seen. Cramer V Effect size is greater than 0.2 in 13 of 21 items. The effect size value of 0.29 indicates a good effect of the intervention in the Experimental

Group. In the remaining items the effect size is 0.17. Odds Ratio is also taken to indicate more likelihood of change in the Experimental Group with some degrees of difference. See Table 4.22.

The neutral option on all items of a single stance was taken by only 3 respondents. Generally, neutral is a random option on items and not a preferred choice which is further reduced at post-test. The neutral is interesting where it is seen that fewer respondents opted for a neutral in the Experimental Group at post-test than the Control Group.

Weighted averages improved at post-test in the Experimental Groups' consistency with the Criterialist stance as assessed with BLTHQ data. Median values show low consistency in the Control Group at post-test while the Experimental Groups show more consistency. A paired t test has a consistency score showing significant difference at post-test in the Experimental Group. (0.002). Factor Analysis improved at post-test in the Experimental Group where a clear Criterialist factor emerges. Generally, it can be said from an examination of the data from the BM and BLTHQ instruments, that there is consistence in a better performance in change in concepts about knowledge in history in the Experimental Group as compared to the Control Group.

The EBI focuses on general epistemic beliefs, therefore a comparison of change in stance from pre-test to post-test was seen independently in the instrument. The data of the EBI supports the findings of more change in the Experimental Group. Of 32 items, 20 showed more change in the Experimental Groups. The Cohen d effect size for percentage difference shows moderate effect of the intervention in the Experimental Group. The Experimental Group showed a disagreement with knowledge as certain at post-test which is interpreted as coherent with findings of the BM and BLTHQ. This result is thus seen as consistent with the result of the BM and BLTHQ analysis.

In an analysis of the 5 independent studies, more and better change in terms of change to a higher stance is seen in 4 of 5 Experimental Groups. This is an important result raising questions of cause. Intervention 1, an experience of a range of suitable examples, was provided to all 10 Groups Control and

Experimental. This was the pedagogical method and it is presented and discussed. Participants in both Control and Experimental Groups changed significantly from pre-test to post-test. Thus, the pedagogical method has been successful in changing epistemic concepts about knowledge in history. Intervention 2, an enhanced perception of agency, was provided to Experimental Groups only in all 5 studies. More and better change is seen with moderate effect sizes in 4 of 5 Experimental Groups and in one Control Group.

The effect of Intervention 2 in the Experimental Group is first seen in the meta-analysis of overall data of the BM. A similar trend is seen in the BLTHQ and EBI data where result of meta-analysis indicates that more items have significant change in the Experimental Group as compared to items of the Control Group.

The number of items being 21 in the BLTHQ Likert Scale and 32 in the EBI, a study wise analysis of difference was difficult. The numbers in cells were small in the single studies as well as unequal in numbers of participants. One single study (Rose - Table 4.21) was analyzed by means of the same statistical test used for meta- analysis of overall data of 5 studies. Though the sample size is smaller, the result is quite consistent and meaningful. Only two items in the Control Group have significant difference in pre-test to post-test responses while in the Experimental Group, 8 items including all 4 Criterialist items show significant difference from pre-test to post-test. This I consider meaningful support to the finding of the BM that 4 out of 5 studies show more positive change in the Experimental Groups. Meta-analysis of overall combined data of the BLTHQ and EBI also supports a conclusion of more and better change in the Experimental Groups.

I, as the tutor, practitioner researcher in all groups, cannot account for any other possible common factor in the 4 Experimental Groups that could be said to be a reason for the difference in epistemic belief change between groups. The tutor, materials, activities, general focus of discussion, school contexts, plan, were the same in all groups. What was consistently maintained as different in the Experimental Groups was the enhanced perception of agency. There were random differences certainly in groups. Size, composition, city, school, qualifications,

experience, etcetera are all discussed as the many varying natural features of the sample in practice contexts. What is known to be common is the enhanced perception of agency in the Experimental Groups alone. This is discussed in the following pages.

5.6 Limitations and Strengths

There are limitations to inference as described in the nature of the constructs of epistemic beliefs and perceptions of agency. The limits of reliability of qualitative assessment of epistemic beliefs with the CBKH category scale and the 3 measures, threats to validity in the quasi experimental design, non-random sample, multiple differences within the sample, field contexts, practitioner research, size of the samples, the limitations of statistical tests conducted are taken into account but cannot be measured and calculated for statistical analysis.

This is a self-funded practitioner research for a degree and there are limitations of access to high powered statistical analysis.

Research in the social sciences is fraught with difficulties and what is claimed as known, takes all such factors into account. As Shadish et al. (2002) warn, and I have found, it is frustrating if after all this effort, non-random selection and weak control in a weak quasi experiment make a more strongly supported conclusion difficult. The practitioner researcher's role, assignment process and unknown basis of the samples present in the 10 workshops are reasons to avoid strong claims. It is reiterated that such limitations are not peculiar to this research but would always be present in field research and research into epistemic beliefs.

To improve the reliability of the qualitative assessment of responses to BM questions, I conducted reasonably rigorous Inter Rater and Internal Evaluation. In so doing, apparently, I limit the subjectivity of the Rater, yet say little about the subjectivity of the rubric (i.e. the categories I create) or the extent to which results can be generalised.

The rubric and the validity of the measures can be defended best by theoretical arguments as I do for the CBKH. To begin, I have used the models of Kuhn and Weinstock (2002) and Maggioni et al. (2004), Maggioni et al. (2009a)

then developed these further. While I argue with and alter the model they employed in the research they conducted, my position is actually consistent with much of what is their position regarding epistemic beliefs in the other writings they have produced as well as others such as Lee and Shemilt (2003). There is, however, difference in emphasis. I set higher standards for a criterialist stance by insisting upon the continued awareness of the constructed nature of knowledge in history and acknowledgement of the historian in the text at the highest level of beliefs. I also increase the spread within the stance to include those who are now aware that there is a way out of the relativist impasse but do not yet know the use of the heuristics well enough. At the highest level of the Sophisticated Integrator, expertise in the use of the historical method is expected as described by these authors at the Criterialist Level. I have separated ways of knowing from the knower and known in the rubric and, this again, I defend as a useful and necessary step. In so doing I separate the essential epistemic aspect of belief from the possibly weak knowledge and skill in evaluating an account with the use of heuristics.

The rubric is also defended as it enables categorising a vast majority of the data. The CBKH was extremely useful in analysing the 166 response sheets with open ended, qualitative responses to the BM questions. It is an important tool to examine large numbers of responses with reasonable reliability. My argument takes into account the responses of my sample and the need to differentiate between these individuals and expert historians. The CBKH rubric is a useful instrument to unravel epistemic beliefs, order these and differentiate between a range of stance at finer grain. More research with trained and tested raters will help to improve reliability. A set of exemplars is provided. There are, at present, nuances of difference but a lot of consensus available in the literature as to what constitutes a naive or sophisticated stance. This is, in the end, a subjective yet necessary decision.

There are limitations to consistency of stance. Weighted averages do not show consistency if consistency with the Criterialist stance is set at 90%. According to the data of the BLTHQ, participants take a dual stance agreeing with both the Criterialist as well as one more, either Borrower or Copier stance. This dual stance

I see as an instrument effect and not coherent with the findings of the BM. VanSledright et al. (2006) describe similar findings in their data as do Maggioni et al. (2009a). The 4 items of the Criterialist stance are seen as hard to disagree with and it is also seen as difficult to disagree, at the same time, with all the remaining 17 in the scale. Therefore, agreement or disagreement with the Borrower and Copier items is taken as adequate in describing the participants' stance. This is consistent with the stance seen in the data of the BM questions and with expectations of change from pre-test to post-test. Factor Analysis does not tell us much about the validity of the BLTHQ measure although more coherence and a clear criterialist stance emerging at post-test is important, in my view. This, I consider a reliable, consistent and arguable result and it provides support to the findings of Maggioni et al. (2004) in their assessment of epistemic beliefs.

There are real strengths in this research. The logical nature of the relationship proposed and its basis in theory of change in paradigms that encourages faith in its effectiveness; validity for inference of research in field contexts; the meaningful nature of the content employed for change in epistemic beliefs, the effectiveness of a sound pedagogical method; conduct of 5 studies in reasonably comparable conditions and the data gathered; the appealing quality of responses to BM questions and the good possibility of assessment that the CBKH provided; the rigorous internal evaluation; the transparent and open reporting; and finally, my, the practitioner researcher's, long years of experience in teaching both children and adults are strong points. The findings, though not as significant as experimental science may require, are certainly persuasive.

Theoretically, the relationship is well worth exploring. Logically too, the relationship makes sense. It would be hard to argue that epistemic concepts can change where there is a perceived lack of agency in the learning context. Behaviour may change at least overtly, concepts probably do not.

5.7 Is the Relationship of a Perception of Agency and Change in Concepts Possible to Research?

I do, in practical terms, conclude that the relationship of a perception of agency with change in concepts can be systematically researched with some rigour as is possible in research into epistemic beliefs. In experimental research, the design becomes a little complicated. All things need to be the same, therefore the pedagogical method such as that of a range of suitable examples has to be specified carefully.

Practically, identifying a concept that needed change, obtaining a suitable sample and relevant measures was possible including finding schools interested in the change. Contextual issues, although challenging and often unplanned and hard to measure, kept occurring but it was still possible to continue with the 5 studies with reasonable success. Data analysis was a large challenge but that, in a properly funded and resourced research, would not be an issue.

A perception of agency is claimed only in the sense that the planned procedures were followed as operationalized and described. Classroom arrangements were maintained. I, as teacher, kept a careful focus on groups to encourage participants to question, agree and disagree with group views, and to reflect upon their own prior ideas as well as the new concepts discussed in the task based discourse. I kept teacher talk to a minimum, and encouraged people to find groups they were comfortable with. It is important not to think of the process as mechanical but in terms of valuing agency, understanding that individuals need opportunities, room to raise their prior ideas, discuss them and make connections with the new. Going back to Kuhn, this research points out, is a need to explore social conditions that have nurtured change in paradigms and relate these to conditions in classrooms.

A focus group session was held with participants of an Experimental Group after the workshops were over. I asked participants to examine the change in their responses to the questionnaires and consider what was it in the way that the workshops were conducted that helped them to change. Interestingly, although

they did not know the intervention or the words ‘a perception of agency’, several identified in various ways, a sense of liberty, and encouragement to think for themselves. Excerpts are added in places to illustrate what participants perceived.

An important point to be made is the need for more qualitative research to give substance to the idea of a perception of agency relating to change in social contexts. I would better employ my qualitative data, adding more excerpts from audio records and field notes in order to improve descriptions in a revision. I had a plan originally to hold focus groups for all workshops to discuss perceptions and change; however, that became practically difficult and only one such session was held.

I used mixed methods and this, I found useful in the research. The responses to open ended questions as qualitative data when interpreted with a rubric made assessment of the large sample possible. Tashakkori and Creswell (2007, p.207-210) propose interconnected qualitative and quantitative components in mixed methods research including starting with a strong mixed methods research question or objective. The focus of my research, a perception of agency as making a difference in epistemic concepts, drove all aspects of the design. The data is mixed as is the analysis, and conclusions are drawn from both qualitative data transformed to quantitative as well as Likert Scale data. As the nature of my qualitative data, type of method, and its analysis are easier to integrate, I faced fewer problems.

The problem with responding to the question if this relationship can be researched on the basis of my data is that I must put forward a strong claim that the cause of the change in my studies was indeed a perception of agency and not some other underlying common factor and this is what is possible to research. I provide the outcome of data analysis as supporting the hypothesis as well as transparent description. I also realise that I do not know what could be the alternative reasons for the consistency seen in most tests, 3 measures and in the results in 4 out of 5 studies. More research is required to respond.

Having said all this, theoretically, there may be limitations to how far the research may be judged as interpretable. An evaluation would be required, as

Meehl (1990) suggests, to make a judgment. The constructs are psychological 'soft areas', the quasi experiments conducted in teacher education classrooms with non-equivalent, non-random groups by a practitioner are quite different from true experiments. This is, however, a good example of practitioner research, and in that sense, the question, is the relationship possible to research, is answered positively in terms of the practical model.

5.8 Quality of Change as seen in the Three Measures and the Use of the CBKH Categories

Change is significant in both groups, Control and Experimental from pre-test to post-test. When the quality of change is analysed, I see that the change is more, one of degree within a stance from a naive to a critical position or to the next stance from Critical Objectivist to a Naive Subjectivist or Critical Subjectivist to a Naive Criterialist. A move from low to high stance is seen but it is not radical but gradual in most cases.

There are similarities and differences between what the data from the three instruments tell us. This is expected due to the indirect approach to epistemic beliefs and contexts of pedagogical practice, in the BLTHQ as well as some instrument effects. The EBI measures general epistemic beliefs from pre-test to post-test and change is seen within these tests. The BM questions are open ended and responses are similarly open and varying.

Importantly, all change is judged as evidence of a perturbation in the individual's beliefs and an attempt to make new meaning, therefore, positive, while 'No Change' is seen as no effect of the Intervention. There are differences and similarities in the quality of data, small differences in degree of change and nature of change in the three sets of data emerging from the 3 measures. The direction of change seen in all three instruments is observed and interpreted towards more sophistication and, again more so, in the Experimental Group. A backward movement towards naive beliefs is seen less often.

The assessment of responses to the BM data shows a coherent pattern. Most participants moved away from an objectivist belief in knowledge of the past,

believing in history as isomorphic with the past or directly obtained from evidence, but did not yet develop the level of knowledge or understanding of heuristics to be categorized as full Criterialists. No one in either group reached Level 7 of the Sophisticated Integrator and, a total of 2 in the Control and 4 of the Experimental out of the total 83, fell in the Critical Criterialist stance at post-test. At pre-test, there was only 1 participant in the entire sample at this stance. 19 moved to, what I term, a Naive Criterialist stance, which is good evidence of development. At the Naive Criterialist stance, the respondent can be said to be aware that there is a way out of subjective helplessness but does not yet know the use of the historical method well enough to be at the Critical level. This stage I argue as the beginning of the Criterialist stance as there is now a coming to light in a sense, an awareness that was not there earlier, that there is a way out of the subjectivist impasse. Both the Naive Criterialist and Critical Criterialist in the CBKH categories fall in, approximately, what Maggioni et al. (2009a) term 'Transition 2' in their model. What they describe as the Criterialist stance in their model is more in keeping with the Sophisticated Integrator in the CBKH categories with additional expectations in the CBKH of epistemic beliefs, expertise, as well as a requirement for the transparent and acknowledged presence of the historian in the text.

Maggioni et al. (2004), Maggioni et al. (2009a) do not report many people in their sample at the Criterialist level in their studies. I considered a wider spread in this category as the research analysis required a differentiation in the sample and a fine measure of change. The CBKH rubric was developed to assess responses in the sample and not a hypothetical sample. The CBKH thus provides a clear reference for epistemic beliefs as assessed. This is described in detail in the Chapter on Research Methodology.

The EBI shows good change but it is more interesting in the small increase with disagreement with knowledge as certain in the Experimental Group. On average, change is observable and more so, significantly, in the Experimental Group and in this sense, the instrument has provided good support. Analysis of the data from the EBI supports the pattern of more change in the Experimental Group in 21 out of 32 items with higher Chi Square values. A moderate effect size

of 0.6 in favour of the Experimental Group is seen, which is somewhat more than the 0.51 effect size seen in BLTHQ data.

The participants' awareness of the historical method was enhanced by the experience of the Intervention and interestingly, they were persuaded towards its effectiveness in selecting between knowledge claims yet they could not securely describe the way in which this could be done. Longer periods of teaching, and more experience in the use of heuristics is recommended. Subjectivist concerns reflected in statements such as ‘all are good’ were happily less likely to be seen in post-test narrative responses to BM questions.

As expected, there is more clarity about change in the narrative responses to the Boscolo and Mason open ended questions. Between the two sets of data, the qualitative analysis of the responses to the BM questions suggests a more coherent move towards sophistication whilst in BLTHQ data the same participants appear to take a dual stance. This can be attributed to the instrument but I also think that as the items target pedagogical practice albeit underpinned by epistemic beliefs, the Intervention would need to overcome both habits of thinking as well as practice, which is again not easy to do.

The data of the BLTHQ is interesting. There is a pattern in change from pre-test to post-test and between groups. Analysing cross tables of each item e.g. Table 4.8 illustrates this. Broadly, the participants in the Control Groups increased in support of the use of the historical method as well as objective knowledge whilst in the Experimental Group, more participants increased in their awareness of the need for method as well as in the constructed nature of knowledge. It is hard to decipher in epistemic terms which group grew in sophistication unless one recognizes the order of the scale of epistemic beliefs. In these frameworks, the Subjectivist stance is in a higher order than the Objectivist stance and the Criterialist is on the top of the scale. Therefore, it can be said that the members of the Experimental Groups grew towards more sophistication in comparison with the Control Groups.

The dual stance seen here was also observed in the data of the authors' studies as well except in the case of experts.[45] This could be an instrument effect in terms of the construction and balance of the items or the pedagogical beliefs that the questions are framed in. VanSledright et al. (2006, p.218) say they are unclear about the answer as to why participants appear to take a dual stance.

As stated above, participants agreed with the Criterialist stance but also agreed with either the Objectivist or Subjectivist stance. No participant, not even the few at a more advanced position as seen in BM data, agreed only with the 4 Criterialist items and disagreed with all 17 other items as experts may be expected to generally do. This is clearly something a revised version of the BLTHQ should address. I have taken this into account and worked around it.

Interestingly, participants do not appear to hold a dual stance in the responses to the more theoretical Boscolo & Mason questions even though there is some evidence of occasional contradiction within responses to the 6 questions. It is perhaps in the nature of epistemic ideas that positions may be exploratory and tentative at times as discussed in the review of literature. More research is needed.

Keeping the seeming desirability of the Criterialist stance in mind, I interpret the dual stance differently. The generally across the board agreement with Criterialist items in the post-test is seen as positive and in keeping with sophistication and then ignored. Agreement with Borrower items is seen as more coherent with instruction, a more sophisticated move towards an awareness of knowledge in history as constructed, as seen in the Experimental Groups. Agreement with Objectivist items is logically seen as a move towards a more Naive Objectivist stance as in the Control Groups.

Considering the framing of epistemic beliefs, Maggioni et al. (2009a, p.208) wonder if the journey towards expertise implies, at a certain point, the adoption of the Borrower stance. I, too, tentatively consider that the Subjectivist stance may be a necessary interim position between the Objectivist and Criterialist stance without which the importance of the use of the historical method to judge between two explanations in history may not seem meaningful. This is the argument I

[45] See Maggioni et al. (2009)

employ in the order of epistemic beliefs in the CBKH scale and, in this sense, there is coherence between the findings of the two instruments. The move towards agreement with a Borrower stance in the Experimental Group is understood in this light.

It could be said that the Criterialist items are hard to disagree with. There is a saying in Urdu that an elephant has different teeth to show off and different to eat with. The straightforward conclusion would be that the Criterialist items had a socially desirable value, therefore most everyone agreed with them. In the post-test, however, whilst the participants of the Experimental Group showed overall, as well as more coherent, change towards the higher stance, also displayed a more independent and thoughtful stance. They registered a balance of change to agreement, disagreement, and No Change in the Criterialist items. The Control Group showed less change overall, more agreement with the Objectivist stance, and a somewhat larger move to the socially desirable Criterialist stance. The awareness of the role of the knower in knowledge construction grows more in the Experimental Groups. Data of the BM responses shows that most people moved from an Objectivist or a Naive Subjectivist stance to a Critical Subjectivist or higher stance in the Experimental Groups.

A key objective of instruction in the workshops was to develop epistemic thinking towards greater sophistication. Direct instruction did not take place. At no point were the words epistemology, epistemological thinking or beliefs mentioned or discussed in direct terms. Historical events and conflicting accounts as well as sources were examined and evidence was studied with the use of heuristics. Those were the tasks carried out and opportunities for epistemic thinking and talk were provided. Therefore, the change towards a more Borrower stance in the Experimental Groups is seen as more coherent with instruction. Participants could only advance in limited ways as there was not enough time for a more concerted experience of historiography. Such are the constraints of research in the field!

A key point is that most people did not fall in the Criterialist stance in the pre-test in BM data and only some moved towards a Criterialist stance at post-test. Now that may be taken as contradicting the BLTHQ data of large numbers

agreeing with the Criterialist stance. If, however, an account is taken of the nature of the question in the BM being dependent on demonstrating knowledge and understanding of the use of heuristics, it is possible to understand the difference. In the BM questions, respondents cannot just take a socially desirable stance. They have to demonstrate they know how to judge between accounts, which is more challenging than selecting what sounds like the right thing to say. Therefore, I trust the BM data and count agreement with Criterialist items in the BLTHQ as positive and necessary. If respondents had not agreed with Criterialist items or agreed both with the Copier and Borrower items that would have been a cause for concern. Happily they did not. The findings in this research are arguably sound. I consider that responses by participants are thoughtful, display coherent patterns, and there is consistency in results of all three instruments. Findings of my research deepen, and are in line with prior findings.

In moving from a Copier or Objectivist stance, the next stage is a Borrower or Subjectivist stance. The Experimental Groups took that direction while the Control Groups moved back to agree with an Objectivist stance. The Experimental Group in this light grew in sophistication whereas the Control Groups moved to a more naive stance. I have ethical concerns about this and want to go back and complete what I began as well as provide equal opportunity to Control Groups to change with an enhanced perception of agency.

Results largely show that the history focused questions of the BM did work to unravel history related epistemic beliefs while the BLTHQ did access the learning and teaching history beliefs held by teachers. Taken together, the two instruments provide a good description of participants' epistemic beliefs about knowledge in history. Maggioni et al. (2009a, p.207) report that most teachers fully completed the questionnaire, which is what I found as well. The neutral option was not taken as a choice and the change in stance found is coherent with expectations. Broadly, the stance seen and change within it is logical and coherent in all groups in the pre-test with a change that is towards more and better development at post-test in the Experimental Groups. This result-as-coherent is also interpreted as support for measurement and analysis.

Combining quantitative Likert Scales with the qualitative Boscolo & Mason Questionnaire was useful in the range of information that could be obtained. Analysis was time consuming as a set of categories to analyze the responses to the open ended questions became necessary and had to be developed. Developing the categories, the CBKH was a challenging enterprise but well worth the time. In the end, the rubric made the analysis with repeated and persistent effort, more reliable. The Inter Rater agreement and Internal Evaluation does suggest that assessment became more consistent with efforts at rigour and repeated trials. I had nagging concerns that reading the responses so many times and coming up with varying results might make the scores I awarded doubtful but the analysis of data has brought up patterns that validate the assessment to a large extent. Pre-test data appears to be similar in many ways in both groups as indeed it does in the BLTHQ and EBI responses. Post-tests show consistent patterns in change and more and better change in the Experimental Groups. The instruments I selected were suitable as described as they assess epistemic beliefs, can be used for a pre-test and post-test measurement of change, and in combination, there is some variance in the construct measured to provide a rich description of stance.

While truth claims about epistemic beliefs need to be tentative, there is enough data in 3 measures over 166 response sheets to give some substance to the argument that beliefs were measured and change was assessed.

5.9 Qualitative and Quantitative Analysis and Interpretations

Mixing methods was liberating and empowering. I was able to select instruments and methods of analysis from both camps and obtain good evidence. The open ended Boscolo & Mason questions provided very useful and illuminating qualitative responses as data in the sample of 83 participants. Obtaining scores for each respondent made a systematic analysis possible of the 166 response sheets. The CBKH categories developed for this research are a useful tool to order and categorise large numbers of responses for an epistemic stance. The process of Inter Rater assessment and Internal Evaluation was necessary to improve

reliability. The Inter Rater exercise was not fruitful in producing trained Raters of a good level as is explained, but the Internal Evaluation did help to improve the quality of assessment to overall 82%. Doing 5 separate studies helped in raising numbers and more than that, a comparison of studies did support drawing conclusions.

I can agree with Limon (2002, p.277-285) that epistemic concepts can be implicit in narrative and can be inferred. At the same time, the other instrument, the BLTHQ Likert Scale Maggioni et al. (2009a)was employed. As described, this has statements about pedagogical practices underpinned by epistemic beliefs and conclusions about beliefs were drawn through a quantitative analysis. A broad degree of concurrence in findings of change in both instruments albeit with differences is argued. The availability of the BLTHQ was timely and useful. Data from this study may help in developing the instrument further.

This is a mixed method research in somewhat hidden ways. The long and arduous analysis, repeated reading and agonizing over meanings and themes in the 166 qualitative response sheets with reflections on epistemic issues was not an easy task. This effort is lost to view as the descriptors awarded are converted into single scores for analysis. The process took over two years, and the development of the CBKH scale, before a satisfactory analysis was completed but it was well worth the effort as the level of reliability of assessment improved with each trial. Patterns in the BM data are logical and consistent as well as theoretically sound, which is the most important evidence in the consideration of reliability. Had the assessment not been sound, there would have been much inconsistency and random differences. All in all, the pattern of findings is persuasive.

Assessment of responses to the CBKH scale needed to be as reliable as possible as indeed all assessment should be. In this research, there was a larger emphasis on reliable judgment as a relationship between two psychological constructs was being explored in experimental ways. Scales of epistemic beliefs are a developing science and, while reliability was improved, it remains a subjective judgment. The Likert Scales, on the other hand, appear a matter of

counting numbers if it were not for the insight that even the Likert Scale score is an effort to quantify an epistemic belief, which is complex.

Data from the Likert Scales was analyzed through quantitative analysis, which, though challenging, seemed fairly straightforward. However, it turns out there is much interpretation involved as the data reveals some aspects clearly and, in others, provides only hints and suggestions. Pulling all the various results of all three instruments together and finding support for arguments and discounting others requires interpretation and meaning making. An example is an analysis of change in the BLTHQ data that finds that respondents from Control and Experimental Groups have gone different ways in the move to the Copier and Borrower stance. This has to be understood in the light of the change of stance seen in the BM data as well as an understanding of how perhaps a perception of agency may have encouraged participants in the Experimental Group to be more independent in selecting stance. The members of the Control Group moved more towards the socially desirable Criterialist stance while those in the Experimental Group changed to a somewhat more balanced Agree and Disagree with the Criterialist stance.

This is very interesting. The participants in the Control Group also moved lower to agreement in the Naive Copier stance, whereas participants in the Experimental Group moved upwards to recognition of the constructed nature of knowledge in history. The numbers can be interpreted and explained in interesting ways.

The open ended questions of the BM related well to the tasks of source evaluation and discussion experienced during the workshops and gave the participants opportunity to show development in their concepts. Their growth in epistemic stance as seen in the measure is coherent with the experience provided to some extent. This is another aspect of interpretation of the scores awarded.

Interestingly, employing the data from each instrument separately to analyze what changed, and did not change, whether the intervention worked to target those particular beliefs or not, can be seen to be an effective strategy for analysis of the BLTHQ and BM, EBI data. The statements of the BLTHQ and BM

questions are different as is the nature of the beliefs that each targets, yet each instrument displays change from pre-test to post-test in its own way and with degrees of difference. This can serve to inform research in the field.

Concluding from all this evidence is a fascinating experience very much like fitting together pieces of a puzzle. To be honest, questions about the status of qualitative or quantitative data or paradigmatic assumptions in analysis were not in the foreground in drawing conclusions. Pragmatic, practical reasoning using numbers, patterns, knowledge such as what could be a socially desirable choice, or what was closer to the content taught, as well as theoretical arguments, helped to draw conclusions.

5.10 Contribution to Knowledge

This research has fortunately proved useful in several ways. A hypothesis of a relationship of a perception of agency in task based group discourse and change in concepts is argued, proposed, and found to be supported in four of five quasi experimental studies. The study also demonstrates how the relationship can be systematically researched. Identifying the relationship opens a small new idea to explore in Kuhn’s theory of paradigm change and generally, the area of conceptual change in education. A perception of agency and change in concepts are discussed in the blurred space of philosophical and psychological arguments about learning where Kuhn's arguments about revolutions in science, Schön's epistemology for reflective practice, counter arguments on convergence of meaning such as those of Newman, and a review of Russell's work on agency and mental development are found. The relationship is briefly introduced and found supported empirically but requires more academic argument for elaboration as well as research.

Testing the relationship in change of other concepts that are problematic in teacher education such as responsibility of pupil behaviour will be interesting. Qualitative research is recommended to better understand the phenomenon of a perception of agency and its effects on learning in social contexts.

Research in epistemic beliefs about knowledge in history is useful in the contexts of curriculum change in history and informs preparations for examination for Cambridge International Examinations in Pakistan. The importance of development of epistemic beliefs to negotiate knowledge claims in a world of multiple media needs increasing focus. The research finds significant support to the question if it is possible to develop epistemic beliefs with adult education in the findings of all 10 groups.

Skemp's principles of experience of a range of suitable examples to change concepts are tried and found most useful in changing concepts, Skemp (1971). I emphasise each aspect, experience, range, suitability, and examples as important in my studies. Significant change in both groups is good evidence for the effectiveness of the strategy.

I developed the Categories of Beliefs about Knowledge in History by modifying Maggioni et al. (2004) Maggioni et al. (2009a) model of the Copier, Borrower, Criterialist Categories. This was necessary as I found that clear descriptions were needed in order to make reliable judgments over large numbers of responses. The new rubric is tested in assessment of 166 response sheets and found useful and reliable. The arguments I have raised and discussed over time with Professor Maggioni have made this possible and I owe her my heartfelt thanks. Maggioni's patient and knowledgeable questions and painstaking reading of my ideas helped to build a better set of categories. Kuhn and Weinstock (2002, p.124), Levels of Epistemological Understanding, were initially studied to develop my own understanding and inform the development. After that, when I saw that my data needed for me to expand and elaborate the categories for more reliable assessment, it was Dr. Liliana Maggioni from the University of Maryland, with her rich experience in the field, who gave of her time and expertise to argue this development. The epistemic arguments for the development of the categories are interesting and the rubric is open for interpretation and practical trial over large data sets. Epistemic belief research would benefit from the description of the combined use of the Boscolo & Mason questions and the assessment of responses with the CBKH categories over large samples.

Carrying out 5 quasi experimental studies in practitioner research in education was instructive. Careful design, countering threats and interpreting the theory in practical contexts was a great learning experience. This should serve as an example of a different way to carry out practitioner research other than case studies. Conducting 5 studies sounds more grand than it is. In schools, teachers often have opportunities to try out a method over time and in varying classrooms. How this can be made more systematic and reliable is useful knowledge.

This is a useful example of practitioner research in teacher education and should be useful and interesting in my field. I have used mixed methods at various levels in measures, analysis and provided transparent and useful descriptions along with the statistics possible. To me personally, mixed methods helped to reflect upon the complexities of paradigms and the implications of stance. I reason that 'personal choice', as a basis of truth claims in educational research, is difficult and lacking coherence with the everyday nature of educational assessment and evaluation. I emphasise the consequences to life chances that truth claims can have in education and try to articulate a stance that is more in keeping with the responsibility of the educational practitioner. I consider 'what works' as a theory to argue that what is known is of value only if it is known how it is known, and to what parameters it can be held to apply. I also give importance to 'what is known to work' in the community of scholars.

In this research the concepts are epistemic beliefs about knowledge in history, which is useful in the contexts of curriculum change in history and informs preparations for examination for Cambridge International Examinations in Pakistan. Teachers' concepts about history knowledge develop with such learning experiences as this excerpt from the focus group illustrates:

'Even Cambridge also gives us that little phrase 'do you agree' that gives you the answer that....Cambridge...that board is also giving them the liberty....there is no question in restricting them to your own opinion, they have the liberty. You should give them the liberty but you should give this aaa idea that they.....it should be proven, they should have valid reason for giving their opinions'

One last point; I share whole heartedly, VanSledright's concern for the teaching of critical reading. He states poignantly that the warrant to teach history in school lies in its opportunity to teach critical reading and that this was for him, 'the only warrant that mattered in the end', VanSledright (2002, p.153). My experience with teachers exploring history accounts and their own epistemic beliefs in the 10 workshops across Pakistan has been that teachers would endorse this point of view as well.

Research in the development of critical thinking is useful in the educational context of Pakistan.

5.11 Alternative Explanations of Cause

Shadish et al. (2002) raise concerns of critics such as Cronbach and Snow (1977) that 'most causal relationships vary across units, settings and times, and so doubt whether there are any constant bivariate causal relationships'. In order to learn if a threat to both was possible and plausible, as argued by Shadish et al. (2002, p.40), I would need to know that the same threat was possible and plausible over the 4 studies out of 5 where the hypothesis is supported in this research. There are a range of broad similarities and fine differences within the persons and settings in the studies, but a common factor would need to be identified. Of the 5 quasi experimental studies conducted in this research, 4 show more change in the Experimental Groups. The workshops were held over five months in three cities and three large school systems of Pakistan. Apart from the Intervention, the common factor could only be myself, the tutor, and my pedagogical method and materials that was working to change concepts. This, to the best of my knowledge, is the enhanced perception of agency as described in the Chapter on Research Methodology. It was the only, conscious, concerted, and organized, difference provided to all 5 Experimental Groups and not the Control Groups. There was, in this 5 month period of the workshops, no other training conducted, which only members of Experimental Groups could have attended.

The other possible and plausible explanation could be, to play the devil's advocate, I read more into the responses and awarded a change category in the Experimental Group that I did not do for the Control Group as many times. That could easily happen in qualitative interpretation if the marking is not blind, systematic, or rigorous. A judgment on epistemic beliefs is difficult and slippery, therefore I used blind marking, developed the category scale and did Inter Rater and Internal Evaluation to improve reliability. However, there is room for improvement.

Another explanation could lie in the statistical analysis. Data from all three measures, all tests in varying degrees and levels of testing, support the alternative hypothesis. Results are consistent in all tests but the reliability and validity of measures and quality of data differ as well as the statistics. Statistical significance in all tests is not to be expected considering the nature of the constructs and the size of the sample. Cronbach (1988) explains that validation tells a party what is strong or weak in their claim. I have tried to do that but conducting an evaluation of validity is difficult for practical reasons described in the document.

Cook et al. (2010, p.114-115) argue that the only defensible warrant for causal assertions is that no alternative explanations are forthcoming from the relevant community of scholars and practitioners concerned with a given subject matter. They suggest that the ultimate warrant for causal inference is not tool based or even truth or logic-based; it is social. It has to do with intense scrutiny by a wide range of knowledgeable others. This is good insight and I would leave it at that.

This is practitioner research using quasi experimental methods and the constraints and possibilities are made transparent. For an alternative explanation other than a perception of agency making a difference, a common cause needs to be identified in the Experimental Groups. That is a complex question to answer. There is coherence and a consistence to the result in many ways which is persuasive.

5.12 Participants Giving Reasons for Change

A most useful, impromptu feedback session was conducted with an Experimental Group 10 days after the last session. I asked the participants to read their responses on both sets of pre-test and post-test to compare the difference. We discussed the change in responses. I then asked them what was it in the sessions, in the way in which they were conducted, that they could recall as having helped them to change. This transcript is most useful in providing some support to the argument that something like a perception of agency was perceived by participants and it was remembered even after the workshops were over. This focus group was in the afternoon at the end of the school day and teachers were busy marking papers and eating a hurried lunch but came good naturedly and sat through the session in a good spirit. They shared the microphone, and the recording fortunately is quite audible, although it is hard to recognize voices partly due to the whirring noise of the fan. The word 'liberty' interestingly, was mentioned in various ways at least five times. The participants of the focus group did not answer straight out but were reflective, thoughtful. Some key themes were, 'you made us think', 'we stirred our minds', 'it depends on the teacher's role', 'not authoritarian, more collaborative', 'depends on how much liberty teacher is giving' 'you gave a chance to students to discuss..... whatever came in their minds'. This recording is valuable as possible evidence of a perception of agency standing out for the participants of the Experimental Group as something that made a difference in changing concepts.

5.13 Final Thoughts

In the end I conclude tentatively that a perception of agency does relate to change in concepts in task based discourse as my hypothesis states. The experience of a perception of agency in my workshops persuaded more people in the Experimental Groups to change their concepts and towards development in epistemic beliefs. Tentatively, this is what the data suggests. As for myself, the

practitioner researcher, I am intrigued by the results and shall try to maintain a perception of agency in task based discourse when working to change concepts in the future.

APPENDIX A1

Case Study 1 and Case Study 2

A Case Study 1

In the video I mention, there is a group of 4 children in an early years classroom (Key Stage 1) working on a map drawing that has to be in keeping with the principles just defined by the teacher of plan views versus side views. The concept of plan view is contra intuitive and learners often draw maps incorporating both plan and side views in the same image. The children appear to have different concepts about this and are unable to find coherence or a meeting ground. One little girl, with strong views, is dominating the exercise and keeps using the eraser to rub off or change what she disapproves of to the frustration of the others. This edgy learning situation is a common feature in schools and one of the reasons perhaps why teachers resist so called group work. Watching the video, I realized that one aspect of the situation could be a perceived lack of agency in the task based discourse which might cause resistance to change. See http://tinyurl.com/mdrpjvd.

A Case Study 2

This is a case of another situation with reference to adult learners where concepts were found to be incongruent amongst a group of teacher educators and a perception of a lack of agency brought about resistance to change. Teacher educators in Pakistan had experience of teaching a course based on a technical rational view of knowledge. Learners were expected to read academic theory and reflect on own practice in its light, generally being critical of own practice. After some years and a change in university, the course was changed radically to one that was premised on practitioner research with its attendant philosophy of knowledge as constructed. Teacher trainers, who were not course developers, were now required to teach the gamut of research process from focusing circles, to selection of criteria and methods to evaluating teaching and learning using

evidence from practice. The trainers, however, whilst teaching its content of teacher research methods, kept up the old approach of technical rationality, one of knowledge as emanating from expert authority. Importantly, when a colleague argued the difference in the two approaches they banded together to resist the idea. Students were instructed that focusing circles were only meant to teach you how to organize your ideas and the requirement of the assignment was to review the literature and reflect upon it. With time however, the change in thinking came about.

APPENDIX-A2

An Information Sheet

Title of Research Project	**:**	Classroom factors that support teacher learning and conceptual change[46]
Principal Researcher	**:**	Lubna Kidwai
Institution	**:**	University of Bradford
Supervisors	**:**	Dr Ivan Reid, Dr George Sheeran
Research Aim	**:**	To investigate a link between classroom process and conceptual change
Charges	**:**	There are no charges for the course.

You are invited to participate in a research project. This is an independent research being conducted by a researcher, Lubna Kidwai, for the purpose of obtaining an award of a PhD from the University of Bradford, UK.

Voluntary Participation

This workshop and participation in research is purely voluntary. You may stop and leave at any time or refuse to answer any question. If you withdraw/leave from the workshop, you will not suffer any personal or professional harm or disadvantage.

In order to help you decide, costs and benefits are described below.

[46] This was the Title at the time of the conduct of research.

Benefits

- A Certificate of completion will be offered to those who complete with full attendance on all three days of the Course.
- You will benefit by working together to understand how key concepts such as cause and significance, and skills such as analysis and reasoning can be taught.
- You will benefit by understanding an ongoing change in the Cambridge History and Culture Syllabus as well as Syllabi around the world.
- You will also benefit by learning good teaching practice in history and being part of a collaborative exercise.
- An opportunity to carry out a project will be offered which you may choose to accept or decline.

Risks/Discomfort

- The training is free of charge, however you will need to pay Rupees 200 for the Certificate. The Certificate is one of attendance and does not require any assessment. Your presence on all days will be enough.
- The training is once a week for 3 weeks which will take your time and effort.
- You will need to fill in several questionnaires which may be time consuming and tiring.
- You will experience some conflict with your existing ideas and the new concepts you will come across.
- You may be requested to permit an audio-record of the proceedings in some sessions which you may accept or refuse as you please.

Procedures to be Followed

- You will be asked to fill out a series of questionnaires and task sheets in the beginning. We will ask for data about yourself such as the class you teach and the numbers of years you have been teaching. This will be recorded with number codes so that you remain anonymous. This data will be stored carefully to maintain confidentiality.
- If you choose to attend the workshop offered, we will record how well you learn, and things that helped you learn.
- At the end of the workshop we may ask you to take part in an interesting feedback session, which we would like to audio record with your permission. If you feel that is not acceptable we will not record the session.

Duration

- The workshops will take place over three days of 3 and a 1/2 hours each.
- The questionnaires/tasks will be filled/carried out during the workshop.

<u>Consent Form</u>

I have read and understood the information given above.
I understand that confidentiality and anonymity will be maintained.
I understand that I will be free to withdraw and leave at any time.
I agree to take part in the research.

Signed : _____
Date : _____

PERSONAL DATA

Name: _____

Gender: Male ☐ Female ☐

Age: 25 – 35 Yrs ☐ 35 – 45 Yrs ☐ 45 – 55 Yrs ☐ 55 & Above ☐

You Teach: Matric ☐ O' Levels ☐ A' Levels ☐

Subjects You Teach: _____

Years You Teach:

Early Years ☐ Primary ☐ Middle School ☐ Senior School ☐

A' Levels ☐

You Teach: Boys ☐ Girls ☐

Qualifications: BA / B Sc ☐ Masters ☐

Any Courses Attended: _____

Years in Teaching: _____

IELTS Score: (if available) _____

APPENDIX-A2a

The History Workshop Certificate

Figure A2a: The History Workshop Certificate

APPENDIX-A2b

History Workshops: Samples in All 5 Studies and in Total for Meta-Analysis

List of Students and Missing Data

(Missing includes those who did not attend on second or third day, did not submit the measure. Those 5 who were teaching Class 8 or Matric are also not included in Missing Data. They were counted out of the analysis even if they attended or submitted the assessment measure.)

S/No	Date	Station	Control		Attended	Missing
			Con	Exp		
1.	6 & 20 January 2011	Rose	14	0	11	3
	7 January & 4 February 2011	‘	0	12	6	6
2.	31 January 2011	Tulip	0	6	6	0
	17 & 31 January 2011	‘	5	0	5	0
3.	27 January & 8 February 2011	Jasmine	0	14	13	1
	26 January & 7 February 2011	‘	15	0	11	4
4.	10-12 May 2011	Bougainvillea	10	0	8	2
	19-21 May 2011	‘	0	12	11	1
5.	16-18 May 2011	Pansy	6	0	5	1
	19-21 May 2011	‘	0	7	7	0
		Total :	**50**	**51**	**83**	**18**

SUMMARY

	Cities	C	E	Total	Missing
a.	Rose	11	6	17	9
b.	Tulip	5	6	11	-
c.	Jasmine	11	13	24	5
d.	Bougainvillea	8	11	19	3
e.	Pansy	5	7	12	1
	G/Total	**40**	**43**	**83**	**18**

APPENDIX-A3

Questions Adapted from those used by Boscolo and Mason

First Name _____ Last Name _____

1. What is History?

2. How do people who write History know about the past they write about?

3. What problems can historians have when they try to understand what happened in the past?

4. Is it possible to explain what happened in different ways?

5. Why?

6. If there are two different explanations, how is it possible to understand which is better?

(Questions adapted from those used by Boscolo and Mason (2001); Bozo, Morra and Pierimarchi, 1989, described in Limon & Mason, 2002: 328)

APPENDIX-A4

Beliefs About Learning and Teaching History

First Name _____ **Last Name** _____

BELIEFS ABOUT LEARNING AND TEACHING HISTORY

1. Students who are good at memorization learn History quickly.
   Strongly Agree…. Agree…. Neutral…. Disagree.... Strongly Disagree

2. Corroborating evidence and identifying sources are important learning strategies in History, but only after mastering the basic facts.
   Strongly Agree…. Agree…. Neutral…. Disagree.... Strongly Disagree

3. In History there is really nothing to understand; the facts speak for themselves.
   Strongly Agree…. Agree…. Neutral…. Disagree.... Strongly Disagree

4. Knowledge of the historical method is fundamental for historians and students alike.
   Strongly Agree…. Agree…. Neutral…. Disagree.... Strongly Disagree

5. Students who know their textbook well will be good at History.
   Strongly Agree…. Agree…. Neutral…. Disagree.... Strongly Disagree

6. To learn History means mainly to study many facts about the past and commit them to memory.
   Strongly Agree…. Agree…. Neutral…. Disagree.... Strongly Disagree

7. Teachers need to avoid giving students conflicting sources, since it makes historical investigation impossible.
   Strongly Agree…. Agree…. Neutral…. Disagree.... Strongly Disagree

8. In learning History, summarizing is more important than comparing.
   Strongly Agree…. Agree…. Neutral…. Disagree.... Strongly Disagree

9. Teachers should not question students' historical opinions, only check that they know the facts.
   Strongly Agree…. Agree…. Neutral…. Disagree.... Strongly Disagree

10. Good general reading and comprehension skills are enough to learn History well.
    Strongly Agree…. Agree…. Neutral…. Disagree.... Strongly Disagree

11. Comparing sources and looking for author subtext are essential components of the process of learning History.
    Strongly Agree…. Agree…. Neutral…. Disagree.... Strongly Disagree

12. Students need to be taught to deal with conflicting evidence.
    Strongly Agree…. Agree…. Neutral…. Disagree.... Strongly Disagree

13. It is fundamental that students are taught to support their reasoning with evidence and ask that History textbook authors do so also.
    Strongly Agree…. Agree…. Neutral…. Disagree.... Strongly Disagree

14. Students who read many History books learn that the past is what the historian makes it to be.
    Strongly Agree…. Agree…. Neutral…. Disagree.... Strongly Disagree

15. Good students know that History is basically a matter of opinion.
Strongly Agree…. Agree…. Neutral…. Disagree.... Strongly Disagree

16. Students need to be aware that History is essentially a matter of interpretation.
Strongly Agree…. Agree…. Neutral…. Disagree.... Strongly Disagree

17. History should be taught like a story: Some things are true, but some others are just a matter of personal opinion.
Strongly Agree…. Agree…. Neutral…. Disagree.... Strongly Disagree

18. In reading a History book, it is more important to pay attention to the perspective of the historian than to his or her reasoning on the evidence discussed.
Strongly Agree…. Agree…. Neutral…. Disagree.... Strongly Disagree

19. Since there is no way to know what really happened in the past, students can believe whatever story they choose.
Strongly Agree…. Agree…. Neutral…. Disagree.... Strongly Disagree

20. Teaching that one historical interpretation is better than another is usually inappropriate.
Strongly Agree…. Agree…. Neutral…. Disagree.... Strongly Disagree

21. Teachers need to make all historical interpretations available and let the students construct their own understanding of them.
Strongly Agree…. Agree…. Neutral…. Disagree.... Strongly Disagree

Explanation of words in italics

Corroborating: Confirming and checking.
Historical method: Methods such as investigating, using sources to develop and evaluate historical accounts.
Conflicting sources: Sources that seem to contradict each other.
Comparing: Comparing sources of evidence.
Author subtext: Underlying, or hidden meaning of the author.

Note:

Scale obtained after written advice from VanSledright, from Maggioni et al. (2009a). The adaptation is only a note with terms in 4 items elaborated.

APPENDIX-A5

Epistemic Beliefs Inventory

First Name ☐ Last Name ☐

Note: In this part, we want you to indicate how strongly you agree or disagree with each of the statements listed below. Please circle the number that best corresponds to the strength of your belief.

1. It bothers me when instructors don't tell students the answers to complicated problems.

Strongly Disagree 1 ☐ 2 ☐ 3 ☐ 4 ☐ 5 ☐ Strongly Agree

2. Truth means different things to different people.

Strongly Disagree 1 ☐ 2 ☐ 3 ☐ 4 ☐ 5 ☐ Strongly Agree

3. Students who learn things quickly are the most successful.

Strongly Disagree 1 ☐ 2 ☐ 3 ☐ 4 ☐ 5 ☐ Strongly Agree

4. People should always obey the law.

Strongly Disagree 1 ☐ 2 ☐ 3 ☐ 4 ☐ 5 ☐ Strongly Agree

5. Some people will never be smart no matter how hard they work.

Strongly Disagree 1 ☐ 2 ☐ 3 ☐ 4 ☐ 5 ☐ Strongly Agree

6. Absolute moral truth does not exist.

Strongly Disagree 1 ☐ 2 ☐ 3 ☐ 4 ☐ 5 ☐ Strongly Agree

7. Parents should teach their children all there is to know about life.

Strongly Disagree 1 ☐ 2 ☐ 3 ☐ 4 ☐ 5 ☐ Strongly Agree

8. Really smart students don't have to work as hard to do well in school.

Strongly Disagree 1 ☐ 2 ☐ 3 ☐ 4 ☐ 5 ☐ Strongly Agree

9. If a person tries too hard to understand a problem, they will most likely end up being confused.

Strongly Disagree 1☐ 2☐ 3☐ 4☐ 5☐ Strongly Agree

10. Too many theories just complicate things.

Strongly Disagree 1☐ 2☐ 3☐ 4☐ 5☐ Strongly Agree

11. The best ideas are often the most simple.

Strongly Disagree 1☐ 2☐ 3☐ 4☐ 5☐ Strongly Agree

12. People can't do too much about how smart they are.

Strongly Disagree 1☐ 2☐ 3☐ 4☐ 5☐ Strongly Agree

13. Instructors should focus on facts instead of theories.

Strongly Disagree 1☐ 2☐ 3☐ 4☐ 5☐ Strongly Agree

14. I like teachers who present several competing theories and let their students decide which is best.

Strongly Disagree 1☐ 2☐ 3☐ 4☐ 5☐ Strongly Agree

15. How well you do in school depends on how smart you are.

Strongly Disagree 1☐ 2☐ 3☐ 4☐ 5☐ Strongly Agree

16. If you don't learn something quickly, you won't ever learn it.

Strongly Disagree 1☐ 2☐ 3☐ 4☐ 5☐ Strongly Agree

17. Some people just have a knack for learning and others don't.

Strongly Disagree 1☐ 2☐ 3☐ 4☐ 5☐ Strongly Agree

18. Things are simpler than most professors would have you believe.

Strongly Disagree 1☐ 2☐ 3☐ 4☐ 5☐ Strongly Agree

19. If two people are arguing about something, at least one of them must be wrong.

Strongly Disagree 1☐ 2☐ 3☐ 4☐ 5☐ Strongly Agree

20. Children should be allowed to question their parents' authority.

Strongly Disagree 1☐ 2☐ 3☐ 4☐ 5☐ Strongly Agree

21. If you haven't understood a chapter the first time through, going back over it won't help.

Strongly Disagree 1 ☐ 2 ☐ 3 ☐ 4 ☐ 5 ☐ Strongly Agree

22. Science is easy to understand because it contains so many facts.

Strongly Disagree 1 ☐ 2 ☐ 3 ☐ 4 ☐ 5 ☐ Strongly Agree

23. The moral rules I live by apply to everyone.

Strongly Disagree 1 ☐ 2 ☐ 3 ☐ 4 ☐ 5 ☐ Strongly Agree

24. The more you know about a topic, the more there is to know.

Strongly Disagree 1 ☐ 2 ☐ 3 ☐ 4 ☐ 5 ☐ Strongly Agree

25. What is true today will be true tomorrow.

Strongly Disagree 1 ☐ 2 ☐ 3 ☐ 4 ☐ 5 ☐ Strongly Agree

26. Smart people are born that way.

Strongly Disagree 1 ☐ 2 ☐ 3 ☐ 4 ☐ 5 ☐ Strongly Agree

27. When someone in authority tells me what to do, I usually do it.

Strongly Disagree 1 ☐ 2 ☐ 3 ☐ 4 ☐ 5 ☐ Strongly Agree

28. People who question authority are trouble makers.

Strongly Disagree 1 ☐ 2 ☐ 3 ☐ 4 ☐ 5 ☐ Strongly Agree

29. Working on a problem with no quick solution is a waste of time.

Strongly Disagree 1 ☐ 2 ☐ 3 ☐ 4 ☐ 5 ☐ Strongly Agree

30. You can study something for years and still not really understand it.

Strongly Disagree 1 ☐ 2 ☐ 3 ☐ 4 ☐ 5 ☐ Strongly Agree

31. Sometimes there are no right answers to life's big problems.

Strongly Disagree 1 ☐ 2 ☐ 3 ☐ 4 ☐ 5 ☐ Strongly Agree

32. Some people are born with special gifts and talents.

Strongly Disagree 1 ☐ 2 ☐ 3 ☐ 4 ☐ 5 ☐ Strongly Agree

SK: Simple Knowledge

Strongly Agree:	1, 10, 11, 13, 18
Strongly Disagree:	24, 30

CK: Certain Knowledge

Strongly Agree:	19, 22, 23, 25
Strongly Disagree:	2, 6, 14, 31

OA: Omniscient Authority

Strongly Disagree:	20
Strongly agree:	4, 7, 27, 28

APPENDIX-A6

EXPLAINING THE RESEARCH DESIGN AND INTERVENTION

(The design is presented here as employed in two studies in order to explain how it is replicated)

Studies	Control Group	Experimental Group	Tests	Data
Study A Context	History workshop for O' Level History teachers related to change in History Syllabus for CIE O' Level Examination towards an assessment of concept of knowledge, evidence and ability to evaluate accounts. Workshop over 3 weeks, 1 day a week	History workshop for O' Level History teachers related to change in History Syllabus for CIE O' Level examination towards an assessment of concept of knowledge, evidence and ability to evaluate accounts. Workshop over 3 weeks, 1 day a week	Pre-Test and Post-Test	Matched 3 Instruments to collect data at Pre-Test and Post-Test. a) BM open ended questions with analysis to CBKH. b) BLTHQ Likert Scale c) EBI Likert Scale
Intervention 1 Necessary for control. Consistent method necessary to avoid threat.	Workshop method Using a range of suitable examples Skemp (1971) of conflicting evidence and accounts	Workshop method Using a range of suitable examples Skemp (1971) of conflicting evidence and accounts	Pre-Test and Post-Test Comparison of Pre-Test and Post-Test performance of all groups.	

Studies	Control Group	Experimental Group	Tests	Data
Intervention 2 (Experimental)		Intervention: Introduction of an enhanced Perception of Agency.	Comparison of Change in stance Control and Experimental Groups	
Study B (Tulip) Context	History workshop for O' Level History teachers related to change in History Syllabus for CIE O' Level Examination towards an assessment of concept of knowledge, evidence and ability to evaluate accounts. Workshop over 3 weeks, 1 day a week	History workshop for O' Level History teachers related to change in History Syllabus for CIE O' Level Examination towards an assessment of concept of knowledge, evidence and ability to evaluate accounts. Workshop over 3 weeks, 1 day a week	Pre-Test and Post-Tests.	Matched 3 Instruments to collect data at Pre-Test and Post-Test. a) BM open ended questions with analysis to CBKH. b) BLTHQ Likert Scale c) EBI Likert Scale

Studies	Control Group	Experimental Group	Tests	Data
Intervention 1 Necessary for control. Consistent method necessary to avoid threat**.**	Workshop method Using a range of suitable examples Skemp (1971) of conflicting evidence and accounts	Workshop method Using a range of suitable examples Skemp (1971) of conflicting evidence and accounts	Pre-Test and Post-Test Comparison of pre-test and post-test performance of all groups.	
Intervention 2		Intervention: Introduction of an enhanced Perception of Agency.	Comparison of change in stance Control and Experimental Groups	

APPENDIX-A7

Evaluation of Qualitative Analysis of Boscolo and Mason's Open Ended Questions to the Rubric, Categories of Beliefs About Knowledge in History (CBKH)

Table A7-1: 1st Internal Evaluation Score Comparison - 5 May 2014

(Random numbers generated by using online Random Number Generator)

Ser	Random No of (1-168)	Group	1st Int Eval Score	Original Score
1.	2	B	6	5
2.	25	A	3	3
3.	27	B	3	4
4.	40	B	4	4
5.	43	B	4	3
6.	44	B	5	4
7.	46	B	4	4
8.	62	B	5	4
9.	63	A	1	1
10.	68	A	4	2
11.	89	B	4	4
12.	90	B	6	6
13.	103	A	4	4
14.	109	B	5	5
15.	114	B	4	2
16.	138	B	3	5
17.	165	B	2	4

Matched Scores: 11 Category Difference: 6

Percentage of Matched Scores: 65% Percentage of mismatch: 35%

Table A7-2: 2nd Internal Evaluation Score Comparison - 11 May 2014

Ser	Random No of (1-168)	Group	2nd Int Eval Score	Original Score
1.	4	B	5	4
2.	18	A	6	3
3.	30	A	4	1
4.	44	A	4	3
5.	49	B	4	1
6.	61	A	3	4
7.	64	A	5	5
8.	71	A	3	2
9.	72	B	2	4
10.	73	B	1	4
11.	79	A	3	1
12.	81	B	3	4
13.	94	B	4	3
14.	110	A	4	4
15.	128	A	4	2
16.	145	A	3	3
17.	153	A	2	4

Matched Scores: 9

Category Difference: 8

Percentage of Matched scores: 53%

Percentage of Mismatch: 47%

Table A7-3: 3rd Internal Evaluation Score Comparison - 24th May 2014

Ser	Random No of (1-168)	Group	3rd Int Eval Score	Original Score	Comparison
1.	5	RC	2	2	Match
2.	9	RC	4	4	M
3.	23	TE	4	4	M
4.	24	TC	4	4	M
5.	31	JE	4	5	CD
6.	41	JE	2	2	M
7.	46	JC	2	2	M
8.	59	RC	3	3	M
9.	64	RE	2	2	M
10.	72	TE	6	6	M
11.	79	TC	2	2	M
12.	85	JE	2	2	M
13.	98	JC	2	2	M
14.	114	BE	4	4	M
15.	116	BC	4	3	CD
16.	123	PC	4	4	M
17.	166	PE	3	2	CD

Matched Scores: 14 Category Difference: 3

Percentage of Matched Scores: 82% Percentage of Mismatch: 17%

Table A7-4: Comparison of Three Internal Evaluations

Internal Eval	Percentage of Match	Percentage of Mismatch
Ist Internal Eval	65%	35%
2nd Internal Eval	53%	47%
3rd Internal Eval	82%	17%

APPENDIX-B

Tables 1B to 27B

Table 1-B: Paired Samples Statistics for Control Group

	Mean	N	Std. Deviation	Std. Error Mean
Pre-test score	3.03	40	0.891	0.141
Post-test score	3.73	40	1.281	0.203

Table 2-B: Paired Samples test for Control Group

	Paired Differences					T	DF	Sig. (2-tailed)
	Mean	Std. Deviation	Std. Error Mean	95% Confidence Interval of the Difference				
				Lower	Upper			
Pre-test score – Post-test score	-0.700	1.067	0.169	-1.041	-0.359	-4.149	39	0.000

Table 3-B: Paired Samples Statistics for Experimental Group

	Mean	N	Std. Deviation	Std. Error Mean
Pretest score	3.02	43	0.963	0.147
Post-test score	3.93	43	1.223	0.186

Table 4-B: Paired Samples test for Experimental Group

	Paired Differences					T	DF	Sig. (2-tailed)
				95% Confidence Interval of the Difference				
	Mean	Std. Deviation	Std. Error Mean	Lower	Upper			
Pre-test score – Post-test score	-0.907	1.130	0.172	-1.255	-.559	-5.263	42	.000

Table 5-B: Mean Ranks of No Change Category in Control and Experimental Group

	Group	N	Mean Rank	Sum of Ranks
No Change	Control	10	6.65	66.50
	Experimental	9	13.72	123.50
	Total	19		

Table 6-B: Comparison of pre and post responses for each city in both Control and Experimental Groups

Cities	**Control Group**		**Experimental Group**	
	Z / P-value	Decision	Z / P-value	Decision
Rose	-2.486/0.013	Significant Difference	-2.041/0.041	Significant Difference
Tulip	-1.841/0.066	No Significant Difference	-2.060/0.039	Significant Difference
Jasmine	-1.66/0.096	No Significant Difference	-2.209/0.027	Significant Difference
Bougainvillea	-6.04/0.546	No Significant Difference	-2.310/0.021	Significant Difference
Pansy	-1.732/0.083	No Significant Difference	-3.8/0.705	No Significant Difference

Table 7-B: Mean Ranks of No Change Category in Control and Experimental Group

	Group	N	Mean Rank	Sum of Ranks
No Change	Control	10	6.65	66.50
	Experimental	9	13.72	123.50
	Total	19		

Table 8-B: Mann Whitney Test Statistics of No Change Category in Control and Experimental Group

	No Change
Mann-Whitney U	11.500
Wilcoxon W	66.500
Z	-2.858
Asymp. Sig. (2-tailed)	0.004

Table 9-B: Mean Ranks of Degree Change Category in Control and Experimental Group

	Group	N	Mean Rank	Sum of Ranks
Degree Change	Control	12	8.33	100.00
	Experimental	15	18.53	278.00
	Total	27		

Table 10-B: Mann Whitney Test Statistics of Degree Change Category in Control and Experimental Group

	Degree Change
Mann-Whitney U	22.000
Wilcoxon W	100.000
Z	-3.630
Asymp. Sig. (2-tailed)	.000

Table 11-B: Mean Ranks of Category Change Category in Control and Experimental Group

	Group	N	Mean Rank	Sum of Ranks
Category Change	Control	18	17.56	316.00
	Experimental	19	20.37	387.00
	Total	37		

Table 12-B: Mann Whitney Test Statistics of Degree Change Category in Control and Experimental Group

	Category Change
Mann-Whitney U	145.000
Wilcoxon W	316.000
Z	-0.829
Asymp. Sig. (2-tailed)	0.407

Table 13-B: Comparison of percentage of pre and post, Low and High stance, in Control and Experimental Groups

Category	Control-Pre test	Control –Post test	Experimental –Pre test	Experimental –Post test
Low	29	13	31	8
%	72.5	32.5	72.1	18.6
High	11	27	12	35
%	27.5	67.5	27.9	81.4
Total	40	40	43	43

Chi - Square = 12.832
DF = 1
P-value = 0.000

Chi - Square = 24.819
DF = 1
P-value = 0.000

Table 14-B: In History there is really nothing to understand; the facts speak for themselves

	Control			Experimental		
	Pre Test	**Post Test**	**Total**	**Pre Test**	**Post Test**	**Total**
Strongly Agree	5	6	11	11	7	18
%	12.5	15.0	13.8	25.6	16.3	20.9
Agree	22	17	39	22	23	45
%	55.0	42.5	48.8	51.2	53.5	52.3
Neutral	3	4	7	2	4	6
%	7.5	10.0	8.8	4.7	9.3	7.0
Disagree	9	5	14	7	5	12
%	22.5	12.5	17.5	16.3	11.6	14.0
Strongly Disagree	1	8	9	1	4	5
%	2.5	20.0	11.3	2.3	9.3	5.8
Total	40	40	80	43	43	86
%	100.0	100.0	100.0	100.0	100.0	100.0

Chi - Square = 7.462
D.F = 4
P-value = 0.113

Chi - Square = 3.711
D.F = 4
P-value = 0.447

Table 15-B: Students who know their textbook well will be good at History

	Control			Experimental		
	Pre Test	**Post Test**	**Total**	**Pre Test**	**Post Test**	**Total**
Strongly Agree	5	5	10	3	1	4
%	12.5	12.5	12.5	7.0	2.3	4.7
Agree	17	17	34	19	29	48
%	42.5	42.5	42.5	44.2	67.4	55.8
Neutral	8	8	16	10	5	15
%	20.0	20.0	20.0	23.3	11.6	17.4
Disagree	9	7	16	10	6	16
%	22.5	17.5	20.0	23.3	14.0	18.6
Strongly Disagree	1	3	4	1	2	3
%	2.5	7.5	5.0	2.3	4.7	3.5
Total	40	40	80	43	43	86
%	100.0	100.0	100.0	100.0	100.0	100.0

Chi - Square = 1.250
D.F = 4
P-value = 0.870

Chi - Square = 6.083
D.F = 4
P-value = 0.193

Table 16-B: To learn History means mainly to study many facts about the past and commit them to memory

	Control			Experimental		
	Pre Test	**Post Test**	**Total**	**Pre Test**	**Post Test**	**Total**
Strongly Agree	9	8	17	9	11	20
%	22.5	20.0	21.3	20.9	25.6	23.3
Agree	11	12	23	14	16	30
%	27.5	30.0	28.8	32.6	37.2	34.9
Neutral	5	5	10	8	5	13
%	12.5	12.5	12.5	18.6	11.6	15.1
Disagree	15	11	26	8	8	16
%	37.5	27.5	32.5	18.6	18.6	18.6
Strongly Disagree	0	4	4	4	3	7
%	0	10.0	5.0	9.3	7.0	8.1
Total	40	40	80	43	43	86
%	100.0	100.0	100.0	100.0%	100.0	100.0

Chi - Square = 4.718
D.F = 4
P-value = 0.318

Chi - Square = 1.168
D.F = 4
P-value = 0.883

Table 17-B: In learning History, summarizing is more important than comparing.

	Control			Experimental		
	Pre Test	**Post Test**	**Total**	**Pre Test**	**Post Test**	**Total**
Strongly Agree	5	12	17	7	10	17
%	12.5	30.0	21.3	16.3	23.3	19.8
Agree	24	16	40	27	24	51
%	60.0	40.0	50.0	62.8	55.8	59.3
Neutral	6	4	10	3	3	6
%	15.0	10.0	12.5	7.0	7.0	7.0
Disagree	5	5	10	5	5	10
%	12.5	12.5	12.5	11.6	11.6	11.6
Strongly Disagree	0	3	3	1	1	2
%	0	7.5	3.8	2.3	2.3	2.3
Total	40	40	80	43	43	86
%	100.0	100.0	100.0	100.0	100.0	100.0

Chi - Square = 7.882
D.F = 4
P-value = 0.096

Chi - Square = 0.706
D.F = 4
P-value = 0.951

Table 18-B: Teachers should not question students' historical opinions only check that they know the facts.

	Control			Experimental		
	Pre Test	**Post Test**	**Total**	**Pre Test**	**Post Test**	**Total**
Strongly Agree	11	9	20	16	13	29
%	27.5	22.5	25.0	37.2	30.2	33.7
Agree	19	21	40	18	22	40
%	47.5	52.5	50.0	41.9	51.2	46.5
Neutral	7	3	10	0	2	2
%	17.5	7.5	12.5	0	4.7	2.3
Disagree	3	6	9	8	3	11
%	7.5	15.0	11.3	18.6	7.0	12.8
Strongly Disagree	0	1	1	1	3	4
%	0	2.5	1.3	2.3	7.0	4.7
Total	40	40	80	43	43	86
%	100.0	100.0	100.0	100.0	100.0	100.0

Chi - Square = 3.900
D.F = 4
P-value = 0.420

Chi - Square = 5.983
D.F = 4
P-value = 0.20

Table 19-B: Good general reading and comprehension skills are enough to learn History well

	Control			Experimental		
	Pre Test	**Post Test**	**Total**	**Pre Test**	**Post Test**	**Total**
Strongly Agree	4	4	8	2	5	7
%	10.0	10.0	10.0	4.7	11.6	8.1
Agree	18	17	35	21	25	46
%	45.0	42.5	43.8	48.8	58.1	53.5
Neutral	7	4	11	5	2	7
%	17.5	10.0	13.8	11.6	4.7	8.1
Disagree	10	13	23	12	11	23
%	25.0	32.5	28.8	27.9	25.6	26.7
Strongly Disagree	1	2	3	3	0	3
%	2.5	5.0	3.8	7.0%	0	3.5
Total	40	40	80	43	43	86
%	100.0	100.0	100.0	100.0	100.0	100.0

Chi - Square = 1.571
D.F = 4
P-value = 0.814

Chi - Square = 5.963
D.F = 4
P-value = 0.202

Borrower Items

Table 20-B: Students who read many History books learn that the past is what the historian makes it to be.

	Control			Experimental		
	Pre Test	**Post Test**	**Total**	**Pre Test**	**Post Test**	**Total**
Strongly Agree	0	3	3	4	0	4
%	0	7.5	3.8	9.3	0	4.7
Agree	5	5	10	7	10	17
%	12.5	12.5	12.5	16.3	23.3	19.8
Neutral	8	6	14	5	3	8
%	20.0	15.0	17.5	11.6	7.0	9.3
Disagree	20	19	39	19	21	40
%	50.0	47.5	48.8	44.2	48.8	46.5
Strongly Disagree	7	7	14	8	9	17
%	17.5	17.5	17.5	18.6	20.9	19.8
Total	40	40	80	43	43	86
%	100.0	100.0	100.0	100.0	100.0	100.0

Chi - Square = 3.311
D.F = 4
P-value = 0.507

Chi - Square = 5.188
D.F = 4
P-value = 0.269

Table 21-B: Good students know that History is basically a matter of opinion.

	Control			Experimental		
	Pre Test	**Post Test**	**Total**	**Pre Test**	**Post Test**	**Total**
Strongly Agree	2	0	2	1	0	1
%	5.0	0	2.5	2.3	0	1.2
Agree	7	7	14	13	9	22
%	17.5	17.5	17.5	30.2	20.9	25.6
Neutral	12	11	23	5	6	11
%	30.0	27.5	28.8	11.6	14.0	12.8
Disagree	15	16	31	18	21	39
%	37.5	40.0	38.8	41.9	48.8	45.3
Strongly Disagree	4	6	10	6	7	13
%	10.0	15.0	12.5	14.0	16.3	15.1
Total	40	40	80	43	43	86
%	100.0	100.0	100.0	100.0	100.0	100.0

Chi - Square = 2.540
D.F = 4
P-value = 0.637

Chi - Square = 2.126
D.F = 4
P-value = 0.713

(Table 14-B also represents the Borrower stance. Most of the respondents Disagree with this Borrower stance at Control and Experimental Group. The Chi square test for both Control and Experimental Group shows no significant difference among pre-test and post-test stance).

Table 22-B: Students need to be aware that history is essentially a matter of interpretation.

	Control			Experimental		
	Pre Test	**Post Test**	**Total**	**Pre Test**	**Post Test**	**Total**
Strongly Agree	0	1	1			
%	0	2.5	1.3			
Agree	5	5	10	4	5	9
%	12.5	12.5	12.5	9.3	11.6	10.5
Neutral	5	5	10	5	4	9
%	12.5	12.5	12.5	11.6	9.3	10.5
Disagree	24	21	45	26	19	45
%	60.0	52.5	56.3	60.5	44.2	52.3
Strongly Disagree	6	8	14	8	15	23
%	15.0	20.0	17.5	18.6	34.9	26.7
Total	40	40	80	43	43	86
%	100.0	100.0	100.0	100.0	100.0	100.0

Chi - Square = 1.486
D.F = 4
P-value = 0.829

Chi - Square = 3.442
D.F = 4
P-value = 0.382

Table 23-B: History should be taught like a story: Some things are true, but some others are just a matter of personal opinion.

	Control			Experimental		
	Pre Test	**Post Test**	**Total**	**Pre Test**	**Post Test**	**Total**
Strongly Agree	0	1	1	2	2	4
%	0	2.5	1.3	4.7	4.7	4.7
Agree	2	4	6	11	9	20
%	5.0	10.0	7.5	25.6	20.9	23.3
Neutral	5	8	13	3	4	7
%	12.5	20.0	16.3	7.0	9.3	8.1
Disagree	24	19	43	19	20	39
%	60.0	47.5	53.8	44.2	46.5	45.3
Strongly Disagree	9	8	17	8	8	16
%	22.5	20.0	21.3	18.6	18.6	18.6
Total	40	40	80	43	43	86
%	100.0	100.0	100.0	100.0	100.0	100.0

Chi - Square = 2.99
D.F = 4
P-value = 0.558

Chi - Square = 0.368
D.F = 4
P-value = 0.985

Table 24-B: In reading a history book, it is more important to pay attention to the perspective of the historian than to his or her reasoning on the evidence discussed.

	Control			Experimental		
	Pre Test	**Post Test**	**Total**	**Pre Test**	**Post Test**	**Total**
Strongly Agree	2	4	6	4	3	7
%	5.0	10.0	7.5	9.3	7.0	8.1
Agree	8	11	19	13	14	27
%	20.0	27.5	23.8	30.2	32.6	31.4
Neutral	14	6	20	10	7	17
%	35.0	15.0	25.0	23.3	16.3	19.8
Disagree	15	14	29	14	13	27
%	37.5	35.0	36.3	32.6	30.2	31.4
Strongly Disagree	1	5	6	2	6	8
%	2.5	12.5	7.5	4.7	14.0	9.3
Total	40	40	80	43	43	86
%	100.0	100.0	100.0	100.0	100.0	100.0

Chi - Square = 7.042
D.F = 4
P-value = 0.134

Chi - Square = 2.746
D.F = 4
P-value = 0.601

Table 25-B: Since there is no way to know what really happened in the past, students can believe whatever story they choose.

	Control			Experimental		
	Pre Test	**Post Test**	**Total**	**Pre Test**	**Post Test**	**Total**
Strongly Agree	6	9	15	7	11	18
%	15.0	22.5	18.8	16.3	25.6	20.9
Agree	19	19	38	24	23	47
%	47.5	47.5	47.5	55.8	53.5	54.7
Neutral	5	2	7	4	2	6
%	12.5	5.0	8.8	9.3	4.7	7.0
Disagree	9	8	17	7	6	13
%	22.5	20.0	21.3	16.3	14.0	15.1
Strongly Disagree	1	2	3	1	1	2
%	2.5	5.0	3.8	2.3	2.3	2.3
Total	40	40	80	43	43	86
%	100.0	100.0	100.0	100.0	100.0	100.0

Chi - Square = 2.278
D.F = 4
P-value = 0.685

Chi - Square = 1.654
D.F = 4
P-value = 0.799

Table 26-B: Teaching that one historical interpretation is better than another is usually inappropriate.

	Control			Experimental		
	Pre Test	**Post Test**	**Total**	**Pre Test**	**Post Test**	**Total**
Strongly Agree	1	1	2	3	4	7
%	2.5	2.5	2.5	7.0	9.3	8.1
Agree	9	18	27	14	15	29
%	22.5	45.0	33.8	32.6	34.9	33.7
Neutral	11	4	15	9	9	18
%	27.5	10.0	18.8	20.9	20.9	20.9
Disagree	15	10	25	13	13	26
%	37.5	25.0	31.3	30.2	30.2	30.2
Strongly Disagree	4	7	11	4	2	6
%	10.0	17.5	13.8	9.3	4.7	7.0
Total	40	40	80	43	43	86
%	100.0	100.0	100.0	100.0	100.0	100.0

Chi - Square = 8.085
D.F = 4
P-value = 0.089

Chi - Square = 0.844
D.F = 4
P-value = 0.932

Table 27-B: Teachers need to make all historical interpretations available and let the students construct their own understanding of them.

	Control			Experimental		
	Pre Test	**Post Test**	**Total**	**Pre Test**	**Post Test**	**Total**
Agree	4	3	7	4	4	8
%	10.0	7.5	8.8	9.3	9.3	9.3
Neutral	5	5	10	3	0	3
%	12.5	12.5	12.5	7.0	0	3.5
Disagree	17	12	29	18	20	38
%	42.5	30.0	36.3	41.9	46.5	44.2
Strongly Disagree	14	20	34	18	19	37
%	35.0	50.0	42.5	41.9	44.2	43.0
Total	40	40	80	43	43	86
%	100.0	100.0	100.0	100.0	100.0	100.0

Chi - Square = 2.064
D.F = 4
P-value = 0.559

Chi - Square = 3.132
D.F = 4
P-value = 0.372

A range is given for the interpretation of estimates of Cramer's V. A value higher than 0.25 indicates very strong relationship, value between 0.15 and 0.25 indicates strong relationship. Any estimates between 0.11 and 0.15 indicate moderate relationship. The estimates between 0.06 and 0.10 show weak relationship and the value between 0.01 and 0.05 shows that the relationship between dependent and independent variable is negligible.

APPENDIX-C

Discussion of Exploratory Factor Analysis

In the Control Group at post-test, Factor 1 explains 21.725% and the cut off value is 0.4, Factor 2 explains 15.89% with the same cut off value as Factor 1. Factor 3 explains 9.719% of variance with a cut off value of 0.5. Factor 4 explains 7.99% of variance with a cut off value of 0.4. The reliability of Factor 1 to Factor 4 is found to be 0.788, 0.720, 0.721 and 0.337 respectively. The percentage of variance and reliability value of Factor 4 and Factors 5, 6, 7 and 8 are very low (see Table 1), that is why I ignored Factors 5, 6, 7 and 8.

Factor 1 appears with all Criterialist items i.e. Item 4: *Knowledge of the historical method is fundamental for historians and students alike,* Item11: *Comparing sources and looking for author subtext are essential components of the process of learning History,* and Item 13: *Students need to be taught to deal with conflicting evidence,* with high positive loading in combination with single Borrower Item i.e. Item 21: *Teachers need to make all historical interpretations available and let the students construct their own understanding of them.* This factor can be taken to reflect a *Criterialist stance* with a combination of Borrower Item i.e. Item 21. The reliability of Factor 1 is found to be 0.778. Factor 2 represents an overall Copier stance and it appears with high positive loading on Item 9 i.e. *Teachers should not question students' historical opinions, only check that they know the facts,* moderate loading on Item 8 i.e. *In learning History, summarizing is more important than comparing,* Item 6 i.e. *To learn History means mainly to study many facts about the past and commit them to memory,* Item 3 i.e. *In History there is really nothing to understand; the facts speak for themselves,* and low positive and negative loading on Item 1 i.e. *Students who are good at memorization learn History quickly*, Item 5 i.e. *Students who know their textbook well will be good at History*, with three Criterialist items i.e. Item 4 i.e. *Knowledge of the historical method is fundamental for historians and students alike,* Item 11 i.e. *Comparing sources and looking for author subtext are essential components of the process of learning History,* Item 13 i.e. *It is fundamental that students are taught to support*

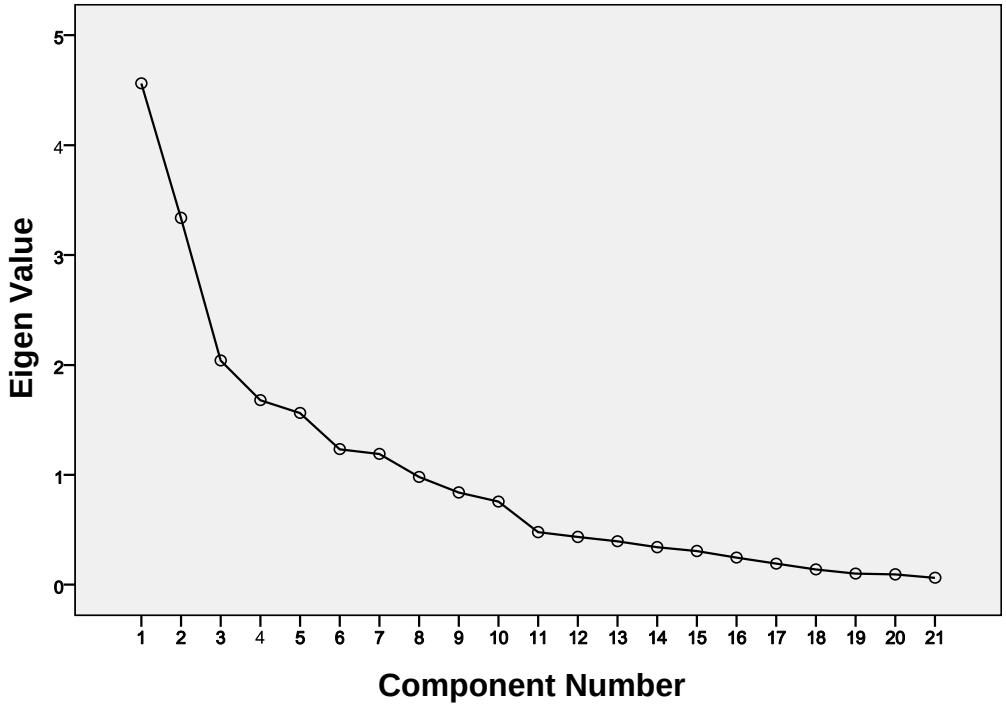
*their reasoning with evidence and ask that History textbook authors do so also. S*o this factor appears to represent a combination of Copier and Criterialist stance. Its cut off value is 0.4 and reliability is 0.720. The cut off value of Factor 3 is 0.5. Factor 3 indicates a combination of a Borrower and Copier stance with high loading on Borrower items i.e. Item 19 i.e. *Since there is no way to know what really happened in the past, students can believe whatever story they choose,* Item 18 i.e. *In reading a History book, it is more important to pay attention to the perspective of the historian than to his or her reasoning on the evidence discussed,* and moderate and low loading on Copier items i.e. Item 5 i.e. *Students who know their textbook well will be good at History,* Item 10 i.e. *Good general reading and comprehension skills are enough to learn History well,* Item 3 i.e. *In History there is really nothing to understand; the facts speak for themselves,* Item 1 i.e. *Students who are good at memorization learn History quickly,* Item 7 i.e. *Teachers need to avoid giving students conflicting sources, since it makes historical investigation impossible,* and a single criterialist item, Item 14 i.e. *Students who read many History books learn that the past is what the historian makes it to be*. Factor 4 also reflects almost the same picture with no certain stance.

In the Experimental Group at post-test, I again considered the first four factors as significant for understanding and interpretation. As stated above, I again opted different cut off values for different factors and considered only those items whose loadings are greater than or equal to these cut off values and ignored all those items whose loadings are less than these cut off values. The percentage of variance and cut off values of all four factors, with their respective reliability values.

Based on loading, Factor 1 emerges completely and wholly a Criterialist stance with very high positive loading on all Criterialist items i.e. Item 4 i.e. *Knowledge of the historical method is fundamental for historians and students alike,* Item 11 i.e. *Comparing sources and looking for author subtext are essential components of the process of learning History,* Item 12 i.e. *Students need to be taught to deal with conflicting evidence,* Item 13 i.e. *It is fundamental that students are taught to support their reasoning with evidence and ask that History textbook authors do so also*. This high loading on Criterialist stance items at post-test differs

from the Control Groups in terms of loading, showing that the respondents' attitude has changed considerably after intervention. Factor 1 of the post-test explains almost 18% of the overall variation in this set of data. I consider this factor can be taken to be a significant factor in my analysis as it clearly points to one single stance that of the Criterialist. The reliability scale for items of Factor 1 is 0.806, which indicates a high consistency in the data. The remaining variation left by Factor 1 is explained in Factor 2. Factor 2 explains 13.57% of variation and its reliability scale value is 0.733, which shows that this factor also contributes a significant role in post-test data. All items in Factors 2 except Item 21 represent the Copier items with very high loading on Item 1 i.e. *Students who are good at memorization learn History quickly,* high loading on Item 6 i.e. *To learn History means mainly to study many facts about the past and commit them to memory,* low loading on Item 17 i.e. *History should be taught like a story: Some things are true, but some others are just a matter of personal opinion,* Item 7 i.e. *Teachers need to avoid giving students conflicting sources, since it makes historical investigation impossible,* Item 2 i.e. *Corroborating evidence and identifying sources are important learning strategies in History, but only after mastering the basic facts,* and Item 3 i.e. *In History there is really nothing to understand; the facts speak for themselves.* Item 21of Borrower stance i.e. *Teachers need to make all historical interpretations available and let the students construct their own understanding of them,* appears with moderate loading. Thus, Factor 2 in the Experimental Group indicates a Copier stance of respondents. Factor 3 appears with high, moderate and low loadings on Borrower, Copier and Criterialist items. This shows the complex behavior of respondents. Factor 4 appears with the same picture as of Factor 3 with no single stance and shows uncertain behaviour of respondents.

Figure 1C: Scree Plot for Combined Data



REFERENCES

Alexander, P. A. and Dochy, F. J. R. C. (1995) Conceptions of Knowledge and Beliefs: A Comparison Across Varying Cultural and Educational Communities. *American Educational Research Journal,* 32 (2)**,** 413-442.

Ashby, R. (2004) Developing a Concept of Historical Evidence: Students' Ideas about Testing Singular Factual Claims. *International Journal of Historical Learning, Teaching and Research,* 4 (2)**,** 1-12.

Ashby, R., Lee, P. J. and Shemilt, D. (2005) Putting Principles into Practice: Teaching and Planning. In: Donovan, M. S. and Bransford, J. D. (Eds.) *How Students Learn: History, Mathematics, and Science in the Classroom.* Washington, D.C.: The National Academies Press, pp. 24-80.

Bain, R. B. (2005) "They Thought the World Was Flat?": Applying the Principles of How People Learn in Teaching Hight School History. In: Donovan, M. S. and Bransford, J., D. (Eds.) *How Students Learn: History, Mathematics, and Science in the Classroom.* Washington, DC: The National Academies Press, pp. 179-213.

Baltas, A. (2007) Background 'Assumptions' and the Grammar of Conceptual Change: Rescuing Kuhn by Means of Wittgenstein. In: Vosniadou, S., et al. (Eds.) *Re-framing the Conceptual Change Approach in Learning and Instruction.* Amsterdam, The Netherlands: Elsevier Ltd, pp. 63-87.

Barton, K. C. (2008) Research on Students' Ideas About History. In: Levstik, L. S. and Tyson, C. A. (Eds.) *Handbook of Research in Social Studies Education.* New York: Routledge, pp. 240-247.

Barton, K. C. and Levstik, L. S. (2004) *Teaching History for the Common Good.* New York; London: Lawrence Erlbaum Associates, Publishers.

Bazeley, P. (2002) Issues in Mixing Qualitative and Quantitative Approaches to Research. In: *1st International Conference - Qualitative Research in Marketing and Management.* University of Economics and Business Administration, Vienna, pp. 1-11.

Belenky, M. F., Clinchy, B. M., Goldberger, N. R. and Tarule, J. M. (1986) *Women's ways of Knowing.* New York: Basic Books, A Member of the Perseus Books Group.

Bendixen, L. D. (2002) A Process Model of Epistemic Belief Change. In: Hofer, B. K. and Pintrich, P. R. (Eds.) *Personal Epistemology : The Psychology of Beliefs About Knowledge and Knowing.* New York; London: Lawrence Erlbaum Associates, Publishers, pp. 191-212.

Boscolo, P. and Mason, L. (2001) Wrting to Learn, Writing to Transfer. In: Tynjala, P., et al. (Eds.) *Writing as a Learning Tool: Integrating Theory and Practice (Studies in Writing).* (Studies in Writing) Vol. 7. Dordrecht; London: Springer (Originally Published by Kluwer Academic Publishers), pp. 83-104.

Bose, S. (2005) The Secession of East Pakistan from Pakistan, 1971, with Two Perspectives on the Causes of Secession. *Economic and Political Weekly,* 40 (41)**,** 4463-4471.

Bransford, J. D. (2000) *How People Learn - Brain, Mind, Experience, and School.* Expanded ed. Washington, D. C. 20418: National Academy Press.

Campbell, D. T. (1957) Factors relevant to the validity of experiments in social settings. *Psychological Bulletin,* 54 (4)**,** 297-312.

Campbell, D. T. and Boruch, R. F. (1975) Making the Case for Randomized Assignment to Treatments by Considering the Alternatives: Six Ways in Which Quasi-Experimental Evaluations In Compensatory Education Tend to Underestimate Effects. In: Bennett, C. A. and Lumsdaine, A. A. (Eds.) *Evaluation and Experiment.* New York: Academic Press Inc., pp. 195-296.

Campbell, D. T. and Stanley, J. C. (1963) *Experimental and Quasi-Experimental Designs for Research.* Chicago: Rand McNally.

Chandler, M. J., Hallett, D. and Sokol, B. W. (2002) Competing Claims About Competing Knowledge Claims. In: Hofer, B. K. and Pintrich, P. R. (Eds.) *Personal Epistemology: The Psychology of Beliefs About Knowledge and Knowing.* New York; London: Lawrence Erlbaum Associates, Publishers, pp. 145-168.

Chi, M. T. H. and Roscoe, R. D. (2002) The Processes and Challenges of Conceptual Change. In: Limon, M. and Mason, L. (Eds.) *Reconsidering Conceptual Change: Issues in Theory and Practice.* Dordrecht; London: Kluwer Academic Publishers, pp. 3-27.

Clinchy, B. M. (2002) Revisiting Women's Ways of Knowing. In: Hofer, B. K. and Pintrich, P. R. (Eds.) *Personal Epistemology: The Psychology of Beliefs About Knowledge and Knowing.* New York; London: Lawrence Erlbaum Associates, Publishers, pp. 63-87.

Cohen, L., Manion, L. and Morrison, K. (2007) *Research Methods in Education.* 6th ed. London: Routledge.

Cook, T. D. and Campbell, D. T. (1979) *Quasi-Experimentation: Design & Analysis Issues for Field Settings.* Chicago: Rand McNally College Publishing.

Cook, T. D., Scriven, M., Coryn, C. L. S. and Evergreen, S. D. H. (2010) Contemporary Thinking About Causation in Evaluation: A Dialogue With Tom Cook and Michael Scriven. *American Journal of Evaluation,* 31 (1)**,** 105-117.

Creswell, J. W. and Plano Clark, V. L. (2011) *Designing and Conducting Mixed Methods Research.* 2nd ed. Los Angeles; London: SAGE Publications Inc.

Cronbach, L. J. (1988) Five Perspectives on Validity Argument. In: Wainer, H. and Braun, H. I. (Eds.) *Test Validity.* Hillsdale, New Jersey: Lawrence Erlbaum Associates, pp. 3-17.

Cronbach, L. J. and Snow, R. E. (1977) *Aptitudes and Instructional Methods: A Handbook for Research on Interactions.* London; New York: Irvington Publishers(Distributed by Wiley).

De Vaus, D. A. (2001) *Research Design in Social Research.* London: SAGE Publications Inc.

diSessa, A. A. (2002) Why "Conceptual Ecology" is a Good Idea. In: Mason, L. and Lucia, M. (Eds.) *Reconsidering Conceptual Change: Issues in Theory and Practice.* Dordrecht; London: Kluwer Academic Publishers, pp. 2-58.
diSessa, A. A. (2006) A History of Conceptual Change Research: Threads and Fault Lines. In: Sawyer, R. K. (Ed.) *The Cambridge Handbook of the Learning Science.* New York: Cambridge University Press, pp. 265-271.
Donovan, M. S. and Bransford, J., D. (2005a) Scientific Inquiry and How People Learn. In: Donovan, M. S. and Bransford, J. D. (Eds.) *How Students Learn: History, Mathematics, and Science in the Classroom.* Washington, D.C.: National Academies Press, pp. 397-416.
Donovan, M. S. and Bransford, J. D. (2005b) Introduction. In: Donovan, M. S. and Bransford, J. D. (Eds.) *How Students Learn: History, Mathematics, and Science in the Classroom.* Washington, D.C.: The National Academies Press, pp. 1-28.
Driver, R. (1983) *The Pupil as Scientist.* Milton Keynes, UK: The Open University Press.
Driver, R. and Easley, J. (1978) Pupils and Paradigms: A Review of Literature Related to Concept Development in Adolescent Science Students. *Studies in Science Education,* 5 (1)**,** 61-84.
Duell, O. K. and Schommer-Aikins, M. (2001) Measures of People's Beliefs About Knowledge and Learning. *Educational Psychology Review,* 13 (4)**,** 420-446.
Ferrari, M. and Elik, N. (2003) Influences on Intentional Conceptual Change. In: Sinatra, G. M. and Pintrich, P. R. (Eds.) *Intentional Conceptual Change.* New York; London: Lawrence Erlbaum Associates, Publishers, pp. 21-54.
Garrison, D. R., Anderson, T. and Archer, W. (2000) Critical inquiry in a text-based environment: Computer conferencing in higher education. *The Internet and Higher Education,* 2 (2-3)**,** 87-105.
Garrison, D. R., Anderson, T. and Archer, W. (2001) Critical Thinking, Cognitive Presence and Computer Conferencing in Distance Education. *American Journal of Distance Education,* 15 (1)**,** 7-23.
Gill, M. G., Ashton, P. T. and Algina, J. (2004) Changing Preservice Teachers' Epistemological Beliefs about Teaching and Learning in Mathematics: An Intervention Study. *Contemporary Educational Psychology,* 29**,** 164-185.
Greene, J. C. (2007) *Mixed Methods in Social Inquiry.* San Francisco, California: Jossey-Bass(John-Wiley and Sons).
Guba, E. G. and Lincoln, Y. S. (1994) *Competing Paradigms in Qualitative Research.* Canada: SAGE Publications Inc.
Hatano, G. and Inagaki, K. (2003) When is Conceptual Change Intended? A Cognitive-Sociocultural View. In: Sinatra, G. M. and Pintrich, P. R. (Eds.) *Intentional Conceptual Change.* New York; London: Lawrence Erlbaum Associates, Publishers, pp. 407-427.
Hattie, J. (2009) *Visible learning: A Synthesis of over 800 meta-analyses relating to achievement.* Abingdon: Routledge(Taylor and Francis Group).
Hewson, P. W. (1992) Conceptual Change in Science Teaching and Teacher Education. In: *Paper presented at a meeting on "Research and Curriculum*

Development in Science Teaching". Madrid, Spain: National Center for Educational Research, Documentation, and Assessment(Ministry for Education and Science), p. 8.
Hoddeson, L. (2007) In the Wake of Thomas Kuhn's Theory of Scientific Revolutions: The Perspective of an Historian of Science. In: Vosniadou, S., et al. (Eds.) *Re-framing the Conceptual Change Approach in Learning and Instruction.* Amsterdam; London; New York: Elsevier Ltd, pp. 25-32.
Hofer, B. K. (2002) Personal Epistemology as a Psychological and Educational Construct: An Introduction. In: Hofer, B. K. and Pintrich, P. R. (Eds.) *Personal Epistemology: A Psychology of Beliefs About Knowledge and Knowing.* London; New York: Lawrence Erlbaum Associates, Publishers, pp. 4-25.
Hofer, B. K. and Bendixen, L. D. (2012) Personal Epistemology: Theory, Research, and Future Directions. In: Harris, K. R., et al. (Eds.) *APA Educational Psychology Handbook.* Vol. 1. American Psychological Association, pp. 227-256.
Howe, K. R. (1988) Against the Quantitative-Qualitative Incompatibility Thesis or Dogmas Die Hard. *Educational Researcher (AERA),* 17 (8)**,** 10-16.
Howe, K. R. and Eisenhart, M. (1990) Standards for Qualitative (and Quantitative) Research: A Prolegomenon. *Educational Researchers (AERA),* 19 (4)**,** 2-9.
Kagan, D. M. (1992) Implications of Research on Teacher Belief. *Educational Psychologist,* 27 (1)**,** 65-90.
Kimmel, A. J. (1988) *Ethics and Values in Applied Social Research.* (Applied Social Research Methods Series) Vol. 12 Newbury Park, California; London: SAGE Publications Inc.
King, P. M. and Kitchener, K. (1994) *Developing Reflective Judgment: Understanding and Promoting Intellectual Growth and Critical Thinking in Adolescents and Adults.* San Francisco: Jossey-Bass.
King, P. M. and Kitchener, K. S. (2002) The Reflective Judgment Model: Twenty Years of Research on Epistemic Cognition. In: Hofer, B. K. and Pintrich, P. R. (Eds.) *Personal Epistemology: The Psychology of Beliefs about Knowledge and Knowing.* New York; London: Lawrence Erlbaum Associates, Publishers, pp. 38-39.
Kloppenburg, J. T. (1996) Pragmatism: An Old Name for Some New Ways of Thinking? *The Journal of American History,* 83 (1)**,** 100-138.
Knowles, M. S., Holton, E. F. and Swanson, R. A. (2005) *The Adult Learner: The Definitive Classic in Adult Education and Human Resource Development.* 6th ed. Amsterdam; London; New York: Elsevier Ltd.
Kuhn, D. and Weinstock, M. (2002) What is Epistemological Thinking and Why Does It Matter? In: Hofer, B. K. and Pintrich, P. R. (Eds.) *Personal Epistemology: The Psychology of Beliefs About Knowledge and Knowing.* New York; London: Lawrence Erlbaum Associates, Publishers, pp. 121-126.
Kuhn, T. S. (1962) *The Structure of Scientific Revolutions.* (International Encyclopedia of Unified Science) Chicago: University of Chicago Press.

Kuhn, T. S. (1970) *The Structure of Scientific Revolutions.* 2nd ed. (International Encyclopedia of Unified Science Foundations of the Unity of Science) Chicago: University of Chicago Press.
Lamb, A. (1992) *Kashmir: A Disputed Legacy 1846-1990.* Karachi: Oxford University Press.
Lave, J. and Wenger, E. (1999) Legitimate Peripheral Participation in Communities of Practice. In: Robert, M. and Carrie, P. (Eds.) *Learning and Knowledge.* London: Paul Chapman Publishing Ltd, pp. 21-35.
Lave, J. and Wenger, E. (2002) Legitimate Peripheral Participation in Communities of Practice. In: Harrison, R., et al. (Eds.) *Supporting Lifelong Learning: Perspectives on Learning.* Vol. 1. Newyork and London: RoutledgeFalmer (The Open University).
LeCompte, M. D. and Goetz, J. P. (1982) Problems of Reliability and Validity in Ethnographic Research. *Review of Educational Research,* 52 (1)**,** 31-60.
Lee, P., Dickinson, A. and Ashby, R. (1993) Progression in Children's Ideas About History. In: *Annual Conference of the British Research Association.* Liverpool, England, United Kingdom: ERIC, pp. 1-26.
Lee, P. J. (2005) Putting Principles into Practice: Understanding History. In: Donovan, M. S. and Bransford, J. D. (Eds.) *How Students Learn: History, Mathematics, and Science in the Classroom.* Washington, DC: The National Academies Press, pp. 31-77.
Lee, P. J. and Shemilt, D. (2003) A Scaffold, not a Cage: Progression and Progression Models in History. *Teaching History (The Historical Association),* 113**,** 13-23.
Limon, M. (2002) Conceptual Change in History. In: Limon, M. and Mason, L. (Eds.) *Reconsidering Conceptual Change: Issues in Theory and Practice.* Dordrecht; London: Kluwer Academic Publishers, pp. 277-285.
London, S. (1996) *Understanding Change: How It Happens and How to Make It Happen.* Available from: http://www.scottlondon.com/reports/change.html (Accessed 13 February 2013).
Loughran, J. and Russell, T. (1997) Meeting Student Teachers on their Own Terms: Experience Precedes Understanding. In: Richardson, V. (Ed.) *Constructivist Teacher Education: Building New Understandings.* London: The Falmer Press, p. 180.
Machamer, P. (2007) Kuhn's Philosophical Successes? In: Vosniadou, S., et al. (Eds.) *Re-framing the Conceptual Change Approach in Learning and Instruction.* (Advances in learning and instruction series) Amsterdam; London; New York: Elsevier Ltd, pp. 35-39.
Maggioni, L., Alexander, P. and VanSledright, B. (2004) At a Crossroads? The Development of Epistemological Beliefs and Historical Thinking. *European Journal of School Psychology,* 2 (1-2)**,** 169-197.
Maggioni, L., Alexander, P. A. and Costich, C. (2007) Teaching History Using Multiple Sources: Exploring Shifts and Pedagogical Implications of Teachers' Epistemic Stance. *Exploring Teachers' Epistemic Stance***,** 6-9.

Maggioni, L., VanSledright, B. and Alexancer, P. A. (2009a) Walking on the Borders: A Measure of Epistemic Cognition in History. *The Journal of Experimental Education,* 77 (3)**,** 187-213.

Maggioni, L., VanSledright, B. and Reddy, K. (2009b) Epistemic Talk: Epistemic Talk in History. In: *The Biennial Confrence of EARLI.* August 2009. pp. 6-207.

Mason, L. (2002) Developing Epistemological Thinking to Foster Conceptual Change in Different Domains. In: Limon, M. and Mason, L. (Eds.) *Reconsidering Conceptual Change: Issues in Theory and Practice.* Dordrecht; London; New York: Kluwer Academic Publishers, pp. 301-335.

Mayer-Smith, J. A. and Mitchell, I. J. (1997) Teaching about Constructivism: Using Approaches Informed by Constructivism. In: Richardson, V. (Ed.) *Constructivist Teacher Education: Building New Understandings.* London: The Falmer Press, p. 129.

McDiarmid, G. W. (1993) Changes in Beliefs about Learners Among Participants in Eleven Teacher Education Programs. In: Gates, J. C. and Peter (Eds.) *Conceptualizing Reflection in Teacher Development.* London: The Falmer Press.

Meehl, P. E. (1990) Why Summaries of Research on Psychological Theories are often Uninterpretable. *Psychological Reports,* 66.

Miles, M. B. and Huberman, A. M. (1994) *Qualitative Data Analysis: An Expanded Sourcebook.* 2nd ed*.* Thousand Oaks, California; London: SAGE Publications Inc.

Mukerjee, M. (2010) *Churchill's Secret War: The British Empire and the Ravaging of India during World War II.* New York: Basic Books.

Newman, S. (1999) *Philosophy and Teacher Education: A Reinterpretation of Donald A. Schön's Epistemology of Reflective Practice.* Aldershot: Ashgate.

O'Grada, C. (2008) The ripple that drowns? Twentieth-Century Famines in China and India as Economic History. *The Economic History Review,* 61 (s1)**,** 5-37.

O'Loughlin (1992) Rethinking Science Education: Beyond Piagetian Constructivism toward a Sociocultural Model of Teaching and Learning. *Journal of Research in Science Teaching,* 29 (8)**,** 791-820.

Paulsen, M. B. and Wells, C. T. (1998) Domain Differences in the Epistemological Beliefs of College Students. *Research in Higher Education,* 39 (4)**,** 365-384.

Piaget, J. (1928) *Judgment and Reasoning in the Child.* London: Routledge & Kegan Paul.

Pintrich, P. R. (2002) Future Challenges and Directions for Theory and Research on Personal Epistemology. In: Hofer, B. K. and Pintrich, P. R. (Eds.) *Personal Epistemology: The Psychology of Beliefs About Knowledge and Knowing.* Vol. 1. New York; London: Lawrence Erlbaum Associates, Publishers, pp. 390-404.

Pintrich, P. R., Marx, R. W. and Boyle, R. A. (1993) Beyond Cold Conceptual Change: The Role of Motivational Beliefs and Classroom Contextual

Factors in the Process of Conceptual Change. *Review of Educational Research,* 63 (2)**,** 167-199.
Posner, G. J., Strike, K. A., Hewson, P. W. and Gertzog, W. A. (1982) Accomodation of a Scientific Conception: Toward a Theory of Conceptual Change. *Science Education,* 66 (2)**,** 211-277.
Reeve, J., Ryan, R., Deci, E. L. and Jang, H. (2008) Understanding and Promoting Autonomous Self Regulation: A Self-Determination Theory Perspective. In: Schunk, D. H. and Zimmerman, B. H. (Eds.) *Motivation and Self-Regulated Learning: Theory, Research and Applications.* London: Lawrence Erlbaum Associates(Taylor Francis Group), pp. 223-244.
Richardson, V. (1997) Constructivist Teaching and Teacher Education: Theory and Practice. In: Richardson, V. (Ed.) *Constructivist Teacher Education: Building New Understandings.* London: The Falmer Press(Taylor and Francis Group), pp. 3-14.
Rogers, A. (2002) Learning and Adult Education. In: Harrison, R., et al. (Eds.) *Supporting Lifelong Learning: Perspectives on Learning.* Vol. 1. London: RoutledgeFalmer (The Open University), pp. 175-178.
Rogers, P. J. (1978) *The New History: Theory into Practice (Teaching of History).* Vol. 44 London: The Historical Association.
Rosenthal, R. and Rosnow, R. L. (2008) *Essentials of Behavioral Research: Methods and Data Analysis.* 3rd ed*.* Boston: McGraw-Hill Humanities/Social Sciences/Languages.
Roth, W.-M. (1999) Authentic School Science: Intellectual Traditions. In: Mccormick, R. and Paechter, C. (Eds.) *Learning and Knowledge.* Paul Chapman Publishing (The Open University), pp. 6-20.
Russell, J. (1995) At Two with Nature: Agency and the Development of Self-World Dualism. In: Bermúdez, J. L., et al. (Eds.) *The Body and the Self.* Cambridge, Massachusetts; London: MIT Press, pp. 127-151.
Russell, J. (1996) *Agency: Its Role in Mental Development.* East Sussex, UK: Erlbaum(UK) Taylor & Francis Lts., Publishers.
Sawyer, R. K. (2006) *The Cambridge Handbook of the Learning Sciences.* Cambridge: Cambridge University Press.
Scheffler, I. (1999) Epistemology and Education. In: Mccormick, R. and Paechter, C. (Eds.) *Learning and Knowledge.* London: Paul Chapman Publishing Ltd, pp. 1-5.
Schommer-Aikens, M. (1990) Effects of Beliefs about the Nature of Knowledge on Comprehension. *Journal of Educational Psychology,* 82 (3)**,** 498-504.
Schön, D. A. (1963) *Displacement of Concepts.* London: Tavistock Publications.
Schön, D. A. (1987) *Educating the Reflective Practitioner: Toward a new design for teaching and learning in the professions.* 1st ed*.* San Francisco, California: Jossey-Bass.
Schraw, G., Bendixen, L. D. and Dunkle, M. E. (2002) Development and Validation of the Epistemic Belief Inventory (EBI). In: Hofer, B. K. and Pintrich, P. R. (Eds.) *Personal Epistemology : The Psychology of Beliefs*

About Knowledge and Knowing. New York; London: Lawrence Erlbaum Associates, Publishers, pp. 261-275.

Schraw, G. J. and Olafson, L. J. (2008) Assessing Teachers' Epistemological and Ontological Worldviews. In: Khine, M. S. (Ed.) *Knowing, Knowledge and Beliefs: Epistemological Studies Across Diverse Cultures.* Dordrecht; London: Springer, pp. 25-29.

Scott, P. H., Asoko, H. M. and Driver, R. H. (1997) Teaching for Conceptual Change: A Review of Strategies. In: Duit, R., et al. (Eds.) *Connecting Research in Physics Education with Teacher Education (An I.C.P.E. Book).* University of Bremen, The International Commission on Physics Education.

Seixas, P. (1996) Conceptualizing the Growth of Historical Understanding. In: Olson, D. R. and Torrance, N. (Eds.) *The Handbook of Education and Human Development.* Oxford: Blackwell Publishers, pp. 765-783.

Shadish, W. R., Clark, M. H. and Steiner, P. R. (2008) Can Nonrandomized Experiments Yield Accurate Answers? A Randomized Experiment Comparing Random and Nonrandom Assignments. *Journal of the American Statistical Association,* 103 (484)**,** 1334-1343.

Shadish, W. R. and Cook, T. D. (2009) The Renaissance of Field Experimentation in Evaluating Interventions. *Annual Review of Psychology,* 60**,** 607-629.

Shadish, W. R., Cook, T. D. and Campbell, D. T. (2002) *Experimental and Quasi-Experimental Design for Generalized Causal Inference.* Belmont, USA: Wadsworth, Cengage Learning.

Shemilt, D. (1980) *History 13 - 16 Evaluation Study.* (Schools Council History 13-16 Project) Edinburgh: Holmes McDougall.

Shemilt, D. (1987) *Adolescent Ideas about Evidence and Methodology in History.* (The History Curriculum of Teachers) London: Falmer Press.

Sinatra, G. M. and Pintrich, P. R. (2003) The Role of Intentions in Conceptual Change Learning. In: Sinatra, G. M. and Pintrich, P. R. (Eds.) *Intentional Conceptual Change.* London: Lawrence Erlbaum Associates, Publishers, pp. 1-16.

Skemp, R. R. (1971) *The Psychology of Learning Mathematics.* London: Penguin Books Ltd.

Smith, J. K. and Heshusius, L. (1986) Closing Down the Conversation: The End of the Quantitative-Qualitative Debate Among Educational Inquirers. *Educational Researcher,* 15 (1)**,** 4-12.

Somekh, B. (2006) Agency, Change and the Generation of Actionable Knowledge. In: Sikes, P. (Ed.) *Action Research: a Methodology Change and Development.* Open University Press, pp. 11-30.

Steiner, P. M., Wroblewski, A. and Cook, T. D. (2009) Randomized Experiments and Quasi-Experimental Designs in Educational Research. In: Ryan, K. E. and Cousins, J. B. (Eds.) *The SAGE International Handbook of Educational Evaluation.* Los Angeles; London: SAGE Publications, Inc, pp. 75-92.

Symonds, J. E. and Gorard, S. (2008) The Death of Mixed Methods: Research Labels and their Casualties. In: *The British Educational Research Association, Annual Conference: September 3-6, 2008.* Heriot Watt University, Edinburgh, pp. 1-17.

Tabak, I. and Weinstock, M. (2008) A Social Exploration of Epistemological Beliefs. In: Khine, M. S. (Ed.) *Knowing, Knowledge and Beliefs: Epistemological Studies Across Diverse Cultures.* Perth, Australia: Springer, pp. 177-195.

Tashakkori, A. and Creswell, J. W. (2007) Exploring the Nature of Research Questions in Mixed Methods Research. *Journal of Mixed Methods Research,* 1 (3)**,** 207-211.

Teddlie, C. and Tashakkori, A. (2009) *Foundations of Mixed Methods Research: Integrating Quantitative and Qualitative approaches in the Social and Behavioral Sciences.* London: SAGE Publications Inc.

Thagard, P. and Zhu, R. (2003) Acupuncture, Incommensurability, and Conceptual Change. In: Sinatra, G. M. and Pintrich, P. R. (Eds.) *Intentional Conceptual Change.* New York; London: Lawrence Erlbaum Associates, Publishers, pp. 79-102.

VanSledright, B. (2002) *In Search of America's Past: Learning to Read History in Elementary School.* New York; London: Teachers College Press, Columbia University.

VanSledright, B. (2009) *The Challenge of Assessing U.S. History Knowledge Growth Among Teachers.* Teaching American History (TAH) Projects. Available from: http://teachinghistory.org/tah-grants/lessons-learned/19432 (Accessed 5 August 2009).

VanSledright, B. (2011) *The Challenge of Rethinking History Education: On Practices, Theories, and Policy.* New York: Routledge.

VanSledright, B. and Afflerbach, P. (2005) Assessing the Status of Historical Sources: An Exploratory Study of Eight US Elementary Students Reading Documents. In: Ashby, R., et al. (Eds.) *Understanding History: Recent Research in History Education, International Review of History Education.* Vol. 4. London; New York: RoutledgeFalmer (Taylor & Francis Group), pp. 1-16.

VanSledright, B., Kelly, T. and Meuwissen, K. (2006) Oh, The Trouble We've Seen: Researching Historical Thinking and Understanding. In: Barton, K. C. (Ed.) *Research Methods in Social Studies Education: Contemporary Issues and Perspectives.* Greenwich, Connecticut: IAP - Information Age Publishing Inc., pp. 207-233.

VanSledright, B. and Reddy, K. (2014) Changing Epistemic Beliefs? An Exploratory Study of Cognition Among Prospective History Teacher. 6, 28-68. Available from http://revistas.udesc.br/index.php/tempo/article/view/21751803061120140 28/3054

VanSledright, B. A. (2014) *Assessing Historical Thinking and Understanding: Innovative Designs for New Standards.* New York & London: Routledge.

Viennot, L. (1979) Spontaneous Reasoning in Elementary Dynamics. *European Journal of Science Education,* 1 (2)**,** 205-221.
Vosniadou, S. (2002) On the Nature of Naive Physics. In: Limon, M. and Mason, L. (Eds.) *Reconsidering Conceptual Change: Issues in Theory and Practice.* Dordrecht; London; New York: Kluwer Academic Publishers, pp. 61-76.
Vosniadou, S. (2003) Exploring the Relationships Between Conceptual Change and Intentional Learning. In: Sinatra, G. M. and Pintrich, P. R. (Eds.) *Intentional Conceptual Change* London: Lawrence Erlbaum Associates, Publishers, pp. 377-406.
Vosniadou, S. (2007a) The Conceptual Change Approach and its Re-Framing. In: Vosniadou, S., et al. (Eds.) *Re-framing the Conceptual Change Approach in Learning and Instruction(Earli).* Amsterdam; London; New York: Elsevier Ltd, pp. 1-30.
Vosniadou, S. (2007b) Personal Epistemology and Personal Change: An Introduction. In: Vosniadou, S., et al. (Eds.) *Re-framing the Conceptual Change Approach in Learning and Instruction.* Amsterdam; London; New York: Elsevier Ltd, pp. 99-104.
Wineburg, S. S. (1991) Historical Problem Solving: A Study of the Cognitive Processes Used in the Evaluation of Documentary and Pictorial Evidence. *Journal of Educational Psychology,* 83 (1)**,** 73-87.
Zeichner, K. M. and Liston, D. P. (1987) Teaching Student Teachers to Reflect. *Harvard Educational Review,* 56 (1)**,** 23-48.